PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12Q 1/68	A1	(11) International Publication Number: WO 93/18186 (43) International Publication Date: 16 September 1993 (16.09.93)

(21) International Application Number: PCT/US93/01775

(22) International Filing Date: 1 March 1993 (01.03.93)

(30) Priority data:

07/846,659	4 March 1992 (04.03.92)	US
07/969,948	30 October 1992 (30.10.92)	US

(71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 300 Lakeside Drive, 22nd Floor, Oakland, CA 94612-3565 (US).

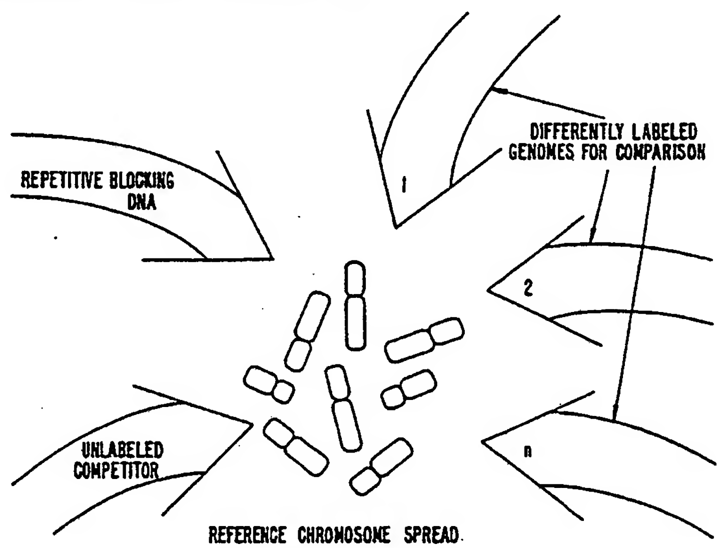
(72) Inventors: PINKEL, Daniel ; 31 Manzanita Court, Walnut Creek, CA 94595 (US). GRAY, Joe, W. ; 1921 11th Street, San Francisco, CA 94116 (US). KALLIONIEMI, Anne ; KALLIONIEMI, Olli, Pekka ; Liljankuja 4, SF-333 00 Tampere (FI). WALDMAN, Frederic ; 206 Edgewood Avenue, San Francisco, CA 94117 (US).

(74) Agent: LAUDER, Leona, L.; 177 Post Street, Suite 800, San Francisco, CA 94108-4731 (US).

(81) Designated States: AU, CA, JP, KR, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: COMPARATIVE GENOMIC HYBRIDIZATION (CGH)



(57) Abstract

Disclosed are new methods comprising the use of *in situ* hybridization to detect abnormal nucleic acid sequence copy numbers in one or more genomes wherein repetitive sequences that bind to multiple loci in a reference chromosome spred are either substantially removed and/or their hybridization signals suppressed. The invention termed Comparative Genomic Hybridization (CGH) provides for methods of determining the relative number of copies of nucleic acid sequences in one or more subject genomes or portions thereof (for example, a tumor cell) as a function of the location of those sequences in a reference genome (for example, a normal human genome). The intensity(ies) of the signals from each labeled subject nucleic acid and/or the differences in the ratios between different signals from the labeled subject nucleic acid sequences are compared to determine the relative copy numbers of the nucleic acid sequences in the one or more subject genomes as a function of position along the reference chromosome spread. Amplifications, duplications and/or deletions in the subject genome(s) can be detected. Also provided is a method of determining the absolute copy numbers of substantially all RNA or DNA sequences in subject cell(s) or cell population(s).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SK	Slovak Republic
CI	Côte d'Ivoire	KZ	Kazakhstan	SN	Senegal
CM	Cameroon	LI	Liechtenstein	SU	Soviet Union
CS	Czechoslovakia	LK	Sri Lanka	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	UA	Ukraine
DK	Denmark	MG	Madagascar	US	United States of America
ES	Spain	ML	Mali	VN	Viet Nam
FI	Finland	MN	Mongolia		

COMPARATIVE GENOMIC HYBRIDIZATION (CGH)

FIELD OF THE INVENTION

This invention relates generally to the field of cytogenetics, and more particularly to the field of molecular cytogenetics. It concerns methods of determining the relative copy numbers of different nucleic acid sequences in a subject cell or cell population and/or comparing the nucleic acid sequence copy numbers of substantially identical sequences in several cells or cell populations as a function of the location of those sequences in a reference genome. For instance, the methods of this invention provide the means to determine the relative number of copies of nucleic acid sequences in one or more subject genomes (for example, the DNA of one tumor cell or a number of cells from a subregion of a solid tumor) or portions thereof as a function of the location of those sequences in a reference genome (for example, a normal human metaphase spread). Further, the invention provides methods of determining the absolute copy number of nucleic acid sequences in a subject cell or cell population.

Although the examples herein concern human cells and the language is primarily directed to human concerns, the concept of this invention is applicable to genomes from any plant or animal. The genomes compared need only be related closely enough to have sufficient substantially identical sequences for a meaningful analysis. For example, a human genome and that of another primate could be compared according to the methods of this invention.

BACKGROUND OF THE INVENTION

Chromosome abnormalities are associated with genetic disorders, degenerative diseases, and exposure to agents known to cause degenerative diseases, particularly cancer, German, "Studying Human Chromosomes Today," American Scientist, 58: 182-201 (1970); Yunis; "The Chromosomal Basis of Human Neoplasia," Science, 221: 227-236 (1983); and German, "Clinical Implication of Chromosome Breakage," in Genetic

Damage in Man Caused by Environmental Agents, Berg, Ed., pgs. 65-86 (Academic Press, New York, 1979). Chromosomal abnormalities can be of several types, including: extra or missing individual chromosomes, extra or missing portions of a chromosome (segmental duplications or deletions), breaks, rings and chromosomal rearrangements, among others. Chromosomal or genetic rearrangements include translocations (transfer of a piece from one chromosome onto another chromosome), dicentrics (chromosomes with two centromeres), inversions (reversal in polarity of a chromosomal segment), insertions, amplifications, and deletions.

Detectable chromosomal abnormalities occur with a frequency of one in every 250 human births. Abnormalities that involve deletions or additions of chromosomal material alter the gene balance of an organism and generally lead to fetal death or to serious mental and physical defects. Down syndrome can be caused by having three copies of chromosome 21 instead of the normal 2. This syndrome is an example of a condition caused by abnormal chromosome number, or aneuploidy. Down syndrome can also be caused by a segmental duplication of a subregion on chromosome 21 (such as, 21q22), which can be present on chromosome 21 or on another chromosome. Edward syndrome (18+), Patau syndrome (13+), Turner syndrome (XO) and Kleinfelter syndrome (XXY) are among the most common numerical aberrations. [Epstein, The Consequences of Chromosome Imbalance: Principles, Mechanisms and Models (Cambridge Univ. Press 1986); Jacobs, Am. J. Epidemiol, 105: 180 (1977); and Lubs et al., Science, 169: 495 (1970).]

Retinoblastoma (del 13q14), Prader-Willis syndrome (del 15q11- q13), Wilm's tumor (del 11p13) and Cri-du-chat syndrome (del 5p) are examples of important disease linked structural aberrations. [Nora and Fraser, Medical Genetics: Principles and Practice, (Lea and Febiger (1989).]

One of the critical endeavors in human medical research is the discovery of genetic abnormalities that are central to adverse health consequences. In many cases, clues to the location of specific genes and/or critical diagnostic markers come from identification of portions of the genome

that are present at abnormal copy numbers. For example, in prenatal diagnosis, as indicated above, extra or missing copies of whole chromosomes are the most frequently occurring genetic lesion. In cancer, deletion or multiplication of copies of whole chromosomes or chromosomal segments, and higher level amplifications of specific regions of the genome, are common occurrences.

Much of such cytogenetic information has come over the last several decades from studies of chromosomes with light microscopy. For the past thirty years cytogeneticists have studied chromosomes in malignant cells to determine sites of recurrent abnormality to glean hints to the location of critical genes. Even though cytogenetic resolution is limited to several megabases by the complex packing of DNA into the chromosomes, this effort has yielded crucial information. Among the strengths of such traditional cytogenetics is the ability to give an overview of an entire genome at one time, permitting recognition of structural abnormalities such as inversions and translocations, as well as deletions, multiplications, and amplifications of whole chromosomes or portions thereof. With the coming of cloning and detailed molecular analysis, recurrent translocation sites have been recognized as involved in the formation of chimeric genes such as the BCR-ABL fusion in chronic myelogeneous leukemia (CML); deletions have been recognized as frequently indicating the location of tumor suppressor genes; and amplifications have been recognized as indicating overexpressed genes.

Conventional procedures for genetic screening and biological dosimetry involve the analysis of karyotypes. A karyotype is the particular chromosome complement of an individual or of a related group of individuals, as defined both by the number and morphology of the chromosomes usually in mitotic metaphase. It include such things as total chromosome number, copy number of individual chromosome types (e.g., the number of copies of chromosome X), and chromosomal morphology, e.g., as measured by length, centromeric index, connectedness, or the like. Karyotypes are conventionally determined by chemically staining an organism's metaphase,

prophase or otherwise condensed (for example, by premature chromosome condensation) chromosomes. Condensed chromosomes are used because, until recently, it has not been possible to visualize interphase chromosomes due to their dispersed condition and the lack of visible boundaries between them in the cell nucleus.

A number of cytological techniques based upon chemical stains have been developed which produce longitudinal patterns on condensed chromosomes, generally referred to as bands. The banding pattern of each chromosome within an organism usually permits unambiguous identification of each chromosome type [Latt, "Optical Studies of Metaphase Chromosome Organization," Annual Review of Biophysics and Bioengineering, 5: 1-37 (1976)].

Unfortunately, such conventional banding analysis requires cell culturing and preparation of high quality metaphase spreads, which is time consuming and labor intensive, and frequently difficult or impossible. For example, cells from many tumor types are difficult to culture, and it is not clear that the cultured cells are representative of the original tumor cell population. Fetal cells capable of being cultured, need to be cultured for several weeks to obtain enough metaphase cells for analysis.

Over the past decade, methods of *in situ* hybridization have been developed that permit analysis of intact cell nuclei--interphase cytogenetics. Probes for chromosome centromeres, whole chromosomes, and chromosomal segments down to the size of genes, have been developed. With the use of such probes, the presence or absence of specific abnormalities can be very efficiently determined; however, it is tedious to test for numerous possible abnormalities or to survey to discover new regions of the genome that are altered in a disease.

The present invention, Comparative Genomic Hybridization (CGH) [formerly called Copy Ratio Reverse Cytogenetics (CRRC) among other names] provides powerful methods to overcome many of the limitations of existing cytogenetic techniques. When CGH is applied, for example, in

the fields of tumor cytogenetics and prenatal diagnosis, it provides methods to determine whether there are abnormal copy numbers of nucleic acid sequences anywhere in the genome of a subject tumor cell or fetal cell or the genomes from representative cells from a tumor cell population or from a number of fetal cells, without having to prepare condensed chromosome spreads from those cells. Thus, cytogenetic abnormalities involving abnormal copy numbers of nucleic acid sequences, specifically amplifications and/or deletions, can be found by the methods of this invention in the format of an immediate overview of an entire genome or portions thereof. More specifically, CGH provides methods to compare and map the frequency of nucleic acid sequences from one or more subject genomes or portions thereof in relation to a reference genome. It permits the determination of the relative number of copies of nucleic acid sequences in one or more subject genomes (for example, those of tumor cells) as a function of the location of those sequences in a reference genome (for example, that of a normal human cell).

Gene amplification is one of several mechanisms whereby cells can change phenotypic expression when increased amounts of specific proteins are required, for example, during development [Spradling and Mahowald, PNAS (USA), 77: 1096-1100 (1980); Glover et al., PNAS (USA), 79: 2947-2951 (1982)], or during an environmental challenge when increased amounts of specific proteins can impart resistance to cytotoxic agents [Melera et al., J. Biol. Chem, 255: 7024-7028 (1980); Beach and Palmiter, PNAS (USA, 78: 2110-2114 (1981)].

A major limitation of Southern analysis and related conventional techniques for analysis of gene amplification is that only specific sites are studied leaving the vast majority of the genome unexamined. Conventional cytogenetic studies, on the other hand, provide a broad survey of the genome but provide little information about genes that may be involved in amplification events. However, the procedures of this invention overcome those limitations. This invention can be used to show the normal chromosomal locations of all regions

of a genome that are amplified or deleted wherein the size of the regions that can be detected is limited only by the resolution of the microscopy used and the organization of DNA in condensed chromosomes. Thus, this invention provides among other uses the ability to study gene amplifications and deletions and their roles in tumor development, progression and response to therapy more thoroughly than was possible previously. The methods of CGH are sufficiently rapid and simple that large numbers of subject nucleic acids, for example from many tumors, can be analysed in studies for gene amplification and deletion.

The karyotypic heterogeneity in solid tumors can be extreme. Identification of commonly occurring chromosomal changes by analysis of metaphase spreads is often difficult or impossible using conventional banding analysis because of the complexity of the rearrangements and because of the poor quality of the metaphase preparations. CGH overcomes that limitation in that the tumor nucleic acid can be studied without the requirement of preparing metaphase spreads. Since CGH can probably be performed on single cells by amplifying the nucleic acid therefrom, CGH can be used to investigate the heterogeneity of tumors by studying representative cells from different cell populations of the tumor. Alternatively, CGH of nucleic acid from a tumor extracted in a bulk extraction process from many cells of the tumor can reveal consistencies within the apparent heterogeneity. For example, the same amplified sequences may appear as homogeneously staining regions (HSRs) and/or double minute chromosomes (DMs) in one tumor cell but as an extension of a chromosome arm in another tumor cell. Thus, order from the apparent randomness may be realized by the use of CGH.

Montgomery et al., PNAS (USA), 80: 5724-5728 (September 1983), concerns the hybridization of labeled Cot fractionated DNAs from tumor cell lines (a Cot fraction from which the high copy repeats, low copy repeats and single copy sequences were substantially removed) to metaphase spreads from said tumor cell lines. Basically, Montgomery et al. mapped the positions of nucleic acid sequences from tumor cell

lines that are very highly amplified back to tumor cell line genomes.

Total genomic DNA from one species has been used in in situ hybridization to discriminate in hybrid cells between chromosomes of that species and of a different species on the basis of the signal from the high copy repetitive sequences. [Pinkel et al., PNAS (USA), 83: 2934 (1986); Manuelidis, Hum. Genet., 71: 288 (1985); and Durnam et al., Somatic Cell Molec. Genet., 11: 571 (1985).]

Landegent et al., Hum. Genet., 77: 366-370 (1987), eliminated highly repetitive sequences, like Alu and Kpn fragments, from whole cosmid cloned genomic sequences by blocking the highly repetitive sequences with Cot-1 DNA. The resulting probe was used for in situ hybridization.

European Patent Application Publication No. 430,402 (published June 5, 1991) describes methods and compositions for chromosome-specific painting, that is, methods and compositions for staining chromosomes based upon nucleic acid sequence employing high complexity nucleic acid probes. In general in the chromosome-specific painting methods, repetitive sequences not specific to the targeted nucleic acid sequences are removed from the hybridization mixture and/or their hybridization capacity disabled, often by blocking with unlabeled genomic DNA or with DNA enriched for high copy repetitive sequences as is Cot-1 [commercially available from Bethesda Research Laboratory, Gaithersburg, MD (USA)]. Pinkel et al., PNAS (USA), 85: 9138-9142 (1988) also describes aspects of chromosome-specific painting as well as International Publication No. WO 90/05789 (published May 31, 1990 entitled "In Situ Suppression Hybridization and Uses Therefor").

Chromosome-specific repeat sequence probes and chromosome-specific painting probes can be hybridized in situ to interphase nuclei as well as metaphase spreads and provide information about the genetic state of the individual targeted genomes. A limitation of such hybridizations is that cytogenetic information is only provided from the regions to which the probes bind. Such hybridizations are very useful

for determining if a particular abnormality is present, for example, the deletion of a specific gene or a duplication among other abnormalities, but it is laborious to search for currently unknown abnormalities on a region by region basis.

Other methods of searching for unknown genetic abnormalities similarly require a lot of work. For example, looking for loss of heterozygosity in tumor cells, requires the hybridization of many probes to Southern blots of tumor and normal cell DNA. CGH provides methods to overcome many of the limitations of the existing cytogenetic techniques.

Saint-Ruf et al., Genes, Chromosomes & Cancer, 2: 18-26 (1990) concluded from their studies of breast cancer that although amplification of genetic material is a frequent and probably important event in breast carcinogenesis, that the relevant genes involved in such amplifications remain unknown but do not seem to correspond to the proto-oncogenes commonly considered important in breast cancer.

Since HSRs in tumors are most often not at the site where the unamplified gene is in normal cells, standard cytogenetics does not yield any information that could assist with identification of the gene(s). CGH on the other hand permits mapping them in the normal genome, a major step towards their identification.

Dutrillaux et al., Cancer Genet. Cytogenet., 49: 203-217 (1990) report (at page 203) that "[a]lthough human breast carcinomas are among the most frequent malignant tumors, cytogenetic data remain scarce, probably because of their great variability and of the frequent difficulty of their analysis." In their study of "30 cases with relatively simple karyotypes to determine which anomalies occur the most frequently and, in particular, early during tumor progression" (p. 203), they concluded that "trisomy 1q and monosomy 16q are early chromosomal changes in breast cancer, whereas other deletions and gain of 8q are clearly secondary events." [Abstract, p. 203.] Dutrillaux et al. further state (at page 216) that deletions within tumor suppressor genes "characterize tumor progression of breast cancer."

It is believed that many solid tumors, such as breast cancer, progress from initiation to metastasis through the accumulation of several genetic aberrations. [Smith et al., Breast Cancer Res. Treat., 18 Suppl. 1: S 5-14 (1991); van de Vijver and Nusse, Biochim. Biophys. Acta, 1072: 33-50 (1991); Sato et al., Cancer Res., 50: 7184-7189 (1990).] Such genetic aberrations, as they accumulate, may confer proliferative advantages, genetic instability and the attendant ability to evolve drug resistance rapidly, and enhanced angiogenesis, proteolysis and metastasis. The genetic aberrations may affect either recessive "tumor suppressor genes" or dominantly acting oncogenes. Deletions and recombination leading to loss of heterozygosity (LOH) are believed to play a major role in tumor progression by uncovering mutated tumor suppressor alleles.

Dominantly acting genes associated with human solid tumors typically exert their effect by overexpression or altered expression. Gene amplification is a common mechanism leading to upregulation of gene expression. [Stark et al., Cell, 75: 901-908 (1989).] Evidence from cytogenetic studies indicates that significant amplification occurs in over 50% of human breast cancers. [Saint-Ruf et al., supra.] A variety of oncogenes have been found to be amplified in human malignancies. Examples of the amplification of cellular oncogenes in human tumors is shown in Table 1 below.

TABLE 1

Amplified Gene	Tumor	Degree of Amplification	DM or HSR Present
c-myc	Promyelocytic leukemia cell line, HL60	20x	+
	Small-cell lung carcinoma cell lines	5-30x	?
N-myc	Primary neuroblastomas (stages III and IV) and neuroblastoma cell lines	5-1000x	+
	Retinoblastoma cell line and primary tumors	10-200x	+
	Small-cell lung carcinoma cell lines and tumors	50x	+
L-myc	Small-cell lung carcinoma cell lines and tumors	10-20x	?
c-myb	Acute myeloid leukemia	5-10x	?
	Colon carcinoma cell lines	10x	?
c-erbB	Epidermoid carcinoma cell	30x	?
	Primary gliomas		?
c-K-ras-2	Primary carcinomas of lung, colon, bladder, and rectum	4-20x	?
N-ras	Mammary carcinoma cell line	5-10x	?

SOURCE: modified from Varmus, Ann. Rev. Genetics, 18: 553-612 (1984) [cited in Watson et al., Molecular Biology of the Gene (4th ed.; Benjamin/Cummings Publishing Co. 1987)]

Chromosomal deletions involving tumor suppressor genes may play an important role in the development and progression of solid tumors. The retinoblastoma tumor suppressor gene (Rb-1), located in chromosome 13q14, is the most extensively characterized tumor suppressor gene [Friend et al., Nature, 323: 643 (1986); Lee et al., Science, 235: 1394 (1987); Fung et al., Science, 236: 1657 (1987)]. The Rb-1 gene product, a 105 kDa nuclear phosphoprotein, apparently plays an important role in cell cycle regulation [Lee et al., supra (1987); Howe et al., PNAS (USA), 87: 5883 (1990)]. Altered or lost expression of the Rb protein is caused by inactivation of both gene alleles either through a

point mutation or a chromosomal deletion. Rb-1 gene alterations have been found to be present not only in retinoblastomas [Friend et al., supra (1986); Lee et al., supra (1987); Fung et al., supra (1987)] but also in other malignancies such as osteosarcomas [Friend et al., supra (1986)], small cell lung cancer [Hensel et al., Cancer Res., 50: 3067 (1990); Rygaard et al., Cancer Res., 50: 5312 (1990)] and breast cancer [Lee et al., Science, 241: 218 (1988); T'Ang et al., Science, 242: 263 (1988); Varley et al., Oncogene, 4: 725 (1989)]. Restriction fragment length polymorphism (RFLP) studies have indicated that such tumor types have frequently lost heterozygosity at 13q suggesting that one of the Rb-1 gene alleles has been lost due to a gross chromosomal deletion [Bowcock et al., Am. J. Hum. Genet., 46: 12 (1990)].

The deletion of the short arm of chromosome 3 has been associated with several cancers, for example, small cell lung cancer, renal and ovarian cancers; it has been postulated that one or more putative tumor suppressor genes is or are located in the p region of chromosome 3 (ch. 3p) [Minna et al., Symposia on Quantitative Biology, Vol. L1: 843-853 (SCH Lab 1986); Cohen et al., N. Eng. J. Med., 301: 592-595 (1979); Bergerham et al., Cancer Res., 49: 1390-1396 (1989); Whang-Peng et al., Can. Genet. Cytogenet., 11: 91-106 (1984; and Trent et al., Can. Genet. Cytogenet., 14: 153-161 (1985)].

The above-indicated collection of amplified and deleted genes is far from complete. As the Saint-Ruf et al. study (supra) of oncogene amplification in cells showing cytogenetic evidence of amplification, such as double minutes (DMs) or homogeneously staining regions (HSRs), indicated, the amplified genes were not known oncogenes in most cases. As Dutrillaux et al., supra indicated, "cytogenetic data remains scarce" for "the most frequent malignant tumors"--breast carcinomas.

Discovery of genetic changes involved in the development of solid tumors has proven difficult. Karyotyping is impeded by the low yield of high quality metaphases and the

complex nature of chromosomal changes [Teyssier, J.R., Cancer Genet. Cytogenet., 37: 103 (1989)]. Although molecular genetic studies of isolated tumor DNA have been more successful and permitted detection of common regions of allelic loss, mutation or amplification [Fearon et al., Cell, 61: 759 (1990); Sato et al., Cancer Res., 50: 7184 (1990); Alitalo et al., Adv. Cancer Res., 47: 235 (1986); and Schwab and Amler, Genes Chrom. Cancer., 1: 181 (1990)], such molecular methods are highly focused, targeting one specific gene or chromosome region at a time, and leaving the majority of the genome unexamined.

Thus, a research tool leading to the identification of amplified and deleted genes and providing more cytogenetic data regarding tumors, especially tumor progression and invasiveness is needed in tumor cytogenetics. CGH provides such a molecular cytogenetic research tool.

The ability to survey the whole genome in a single hybridization is a distinct advantage over allelic loss studies by restriction fragment length polymorphism (RFLP) that target only one locus at a time. RFLP is also restricted by the availability and informativeness of polymorphic probes.

CGH facilitates the genetic analysis of tumors in that it provides a copy number karyotype of the entire genome in a single step. Regions of tumor DNA gain and loss are mapped directly onto normal chromosomes. Comparisons of primary tumors with their metastases by CGH should be informative concerning cancer progression. Analogously, other genomes other than those of tumors can be studied by CGH.

SUMMARY OF THE INVENTION

Comparative Genomic Hybridization (CGH) employs the kinetics of in situ hybridization to compare the copy numbers of different DNA or RNA sequences from a sample, or the copy numbers of different DNA or RNA sequences in one sample to the copy numbers of the substantially identical sequences in another sample. In many useful applications of CGH, the DNA or RNA is isolated from a subject cell or cell population. The comparisons can be qualitative or quantitative.

Procedures are described that permit determination of the absolute copy numbers of DNA sequences throughout the genome of a cell or cell population if the absolute copy number is known or determined for one or several sequences. The different sequences are discriminated from each other by the different locations of their binding sites when hybridized to a reference genome, usually metaphase chromosomes but in certain cases interphase nuclei. The copy number information originates from comparisons of the intensities of the hybridization signals among the different locations on the reference genome.

Two representative basic approaches are employed in CGH as illustrated herein for the analysis of subject DNAs. In an example of the first approach, genomic DNA from a subject cell or cell population of cells is isolated, labeled and hybridized to reference chromosomes, usually in metaphase. In an example of the second approach, genomic DNAs from two or more subject cells or cell populations are isolated, differentially labeled, and hybridized to reference chromosomes, usually in metaphase.

The CGH methods of this invention can be qualitative and/or quantitative. A particular utility of CGH is for analysing DNA sequences from subject cell(s) or cell population(s), for example from clinical specimens including tumor and fetal tissues.

An important utility of CGH is to find regions in normal genomes which when altered in sequence copy number contribute to disease, as for example, cancer or birth defects. For example, regions at elevated copy number may contain oncogenes, and regions present at decreased copy number may contain tumor suppressor genes.

A representative CGH method is for comparing copy numbers of different DNA sequences in a subject cell or cell population comprising the steps of:

a) extracting the DNA from the subject cell or from a number of cells of the subject cell population;

b) amplifying said extracted subject DNA, if necessary;

c) labeling the subject DNA;

d) hybridizing said labeled subject DNA _in situ_ to reference metaphase chromosomes after substantially removing from the labeled DNA those repetitive sequences that could bind to multiple loci in the reference metaphase chromosomes, and/or after blocking the binding sites for those repetitive sequences in the reference metaphase chromosomes by prehybridization with appropriate blocking nucleic acids, and/or blocking those repetitive sequences in the labeled DNA by prehybridization with appropriate blocking nucleic acid sequences, and/or including such blocking nucleic acid sequences for said repetitive sequences during said hybridization, wherein the DNA sequences in the labeled subject DNA that bind to single copy sequences in the reference metaphase chromosomes are substantially retained, and those single copy DNA sequences as well as their binding sites in the reference metaphase chromosomes remain substantially unblocked both before and during the hybridization;

e) rendering the bound, labeled DNA sequences visualizable, if necessary;

f) observing and/or measuring the intensity of the signal from the labeled subject DNA sequences as a function of position on the reference metaphase chromosomes; and

g) comparing the copy numbers of different DNA sequences of the subject DNA by comparing the signal intensities at different positions on the reference metaphase chromosomes, wherein the greater the signal intensity at a given position, the greater the copy number of the sequences in the subject DNA that bind at that position. An analogous method can be performed wherein the subject nucleic acid is RNA.

Further, disclosed are methods whrein two or more subject nucleic acids are analysed by CGH. Exemplary methods are those wherein the subject nucleic acids are DNA sequences from a subject cell or cell population. Analogous methods may be performed wherein the subject nucleic acids are RNA. Such an exemplary method is that for comparing copy numbers of

different DNA sequences in one subject cell or cell population relative to copy numbers of substantially identical sequences in another cell or cell population, said method comprising the steps of:

a) extracting the DNA from both of the subject cells or cell populations;

b) amplifying said extracted subject DNAs, if necessary;

c) differentially labeling the subject DNAs;

d) hybridizing said differentially labeled subject DNAs <u>in situ</u> to reference metaphase chromosomes after substantially removing from the labeled DNAs those repetitive sequences that could bind to multiple loci in the reference metaphase chromosomes, and/or after blocking the binding sites for those repetitive sequences in the reference metaphase chromosomes by prehybridization with appropriate blocking nucleic acids, and/or blocking those repetitive sequences in the labeled DNA by prehybridization with appropriate blocking nucleic acid sequences, and/or including such blocking nucleic acid sequences for said repetitive sequences during said hybridization;

e) rendering the bound, differentially labeled DNA sequences visualizable, if necessary;

f) observing and/or measuring the intensities of the signals from each subject DNA, and the relative intensities, as a function of position along the reference metaphase chromosomes; and

g) comparing the relative intensities among different locations along the reference metaphase chromosomes wherein the greater the intensity of the signal at a location due to one subject DNA relative to the intensity of the signal due to the other subject DNA at that location, the greater the copy number of the sequence that binds at that location in the first subject cell or cell population relative to the copy number of the substantially identical sequence in the second subject cell or cell population that binds at that location.

Further disclosed are methods of quantitatively comparing copy numbers of different DNA sequences in one

subject cell or cell population relative to copy numbers of substantially identical sequences in another subject cell or cell population. A representative method is that comprising steps (a) through (e) of the method immediately detailed above and the following steps of:

f. measuring the intensities of the signals from each of the bound subject DNAs and calculating the ratio of the intensities as a function of position along the reference metaphase chromosomes to form a ratio profile; and

g. quantitatively comparing the ratio profile among different locations along the reference metaphase chromosomes, said ratio profile at each location being proportional to the ratio of the copy number of the DNA sequence that bind to that location in the first subject cell or cell population to the copy number of substantially identical sequences in the second cell or cell population.

Said representative methods can further comprise comparing copy numbers of different DNA sequences in more than two subject DNAs wherein the comparing is done pairwise between the signals from each subject DNA.

This invention further discloses methods to determine the ratio of copy numbers of different DNA sequences in one subject cell or cell population to copy numbers of substantially identical sequences in another cell or cell population wherein the steps of (a) through (f) as described above are performed as well as the following steps:

g. determining the average copy number of a calibration sequence in both subject cells or cell populations, said calibration sequence being substantially identical to a single copy sequence in the reference metaphase cells; and

h. normalizing the ratio profile calculated in (f) so that at the calibration position, the ratio profile is equal to the ratio of the average copy numbers determined in (g), the normalized ratio profile at any other location along the reference metaphase chromosomes thereby giving the ratio of the copy numbers of the DNA sequences in the two subject DNAs that bind at that location. That method can be extended

to further subject nucleic acids as for example determining the ratio of copy numbers of DNA sequences in more than two subject DNAs wherein the comparing is done pairwise between signals from each subject DNA.

Further disclosed are methods for comparing copy numbers of different DNA sequences in a test cell or cell population, said method comprising applying steps (a) through (e) of the above-described methods and

f. observing and/or measuring the intensities of the signal from each subject DNA, and the relative intensities, as a function of position along the reference metaphase chromosomes wherein one of the subject cells or cell populations is the test cell or cell population and the other is a normal cell or cell population; and

(g) comparing the relative intensities among different locations along the reference metaphase chromosomes, wherein the greater the relative intensity at a location, the greater the copy number of the sequence in the test cell or cell population that binds to that location, except for sex chromosomes where the comparison needs to take into account the differences in copy numbers of sequences in the sex chromosomes in relation to those on the autosomes in the normal subject cell or cell population.

A related representative method is that for comparing the copy number of different DNA sequences in a test cell or cell population comprising applying steps (a) through (e) of the above described methods wherein one of the subject cells or cell populations is the test cell or cell population, and the other is a standard cell or cell population wherein the copy numbers of the DNA sequences that bind to different positions on the reference metaphase chromosomes is known and steps:

f. measuring the intensities of the signals from each of the bound subject DNAs and calculating the ratio of intensities as a function of position along the reference metaphase chromosomes to form a ratio profile;

g. adjusting the ratio profile at each location along the reference metaphase chromosomes by multiplying the

ratio profile by the known copy number of DNA sequences in the standard cell or cell population that bind there; and

h. comparing the adjusted ratio profiles at different locations along the reference metaphase chromosomes wherein the greater the adjusted ratio profile at a location, the greater the copy number of the DNA sequence in the test cell or cell population that binds there.

Another related representative method is that for determining the ratios of the copy numbers of different DNA sequences in a test cell or cell population, said method comprising applying steps (a) through (f) of the immediately above-described method and the steps of adjusting the ratio profile at each location along the reference metaphase chromosomes by multiplying the ratio profile by the known copy number of sequences that bind there; and calculating the ratio of the copy number of a DNA sequence in the test cell or cell population that binds to one location on the reference metaphase chromosomes to the copy number of a sequence that binds to another location by dividing the adjusted ratio profile at the location of the first sequence by that at the location of the second. Said representative method can be extended to determine the copy number of different DNA sequences in a test cell or cell population wherein steps (a) through (f) as described above are followed and then the following steps of adjusting the ratio profile at each location along the reference metaphase chromosomes by multiplying the ratio profile by the known copy number of DNA sequences in the standard cell or cell population that bind there;

determining the copy number of a calibration sequence in the test cell or cell population that is substantially identical to a single copy sequence in the reference cells; and

normalizing the adjusted ratio profile so that at the location of the calibration sequence on the reference metaphase chromosomes, the normalized, adjusted ratio profile is equal to the copy number of the calibration sequence determined in the above step, the value of the normalized,

adjusted ratio profile at another location then being equal to the copy number of the DNA sequence in the test cell or cell population that binds at that location. That method can be analogously performed wherein two or more calibration sequences are used, and the adjusted ratio profile is normalized to get the best fit to the copy numbers of the ensemble of calibration sequences. Preferably, the copy number of the calibration sequence is determined by in situ hybridization. Those methods can comprise in situ hybridizing probes for more than one calibration position and normalizing to obtain the best fit of the ratio profile to the calibration positions. The standard cell or cell population preferably have normal genomes. In many applications of CGH, the reference metaphase chromosomes are normal.

Further, this invention concerns the use of antenna cell lines. An exemplary method is for detecting amplification of a certain sequence or group of sequences in a subject cell or cell population, comprising essentially steps (a) through (e) of the above-described methods wherein the in situ hybridization is targeted to antenna cells in which the DNA sequence(s) to be tested for is or are amplified, and examining the reference cell for regions that are hybridized significantly more intensely than others, the presence of such regions indicating amplifications of the sequence(s) which are being tested. The chromosomes of said antenna cell lines may be in interphase or in metaphase.

When a single labeled subject nucleic acid is being hybridized, or if multiple labeled subject nucleic acids are hybridized sequentially, it is important that the binding sites on the reference genome not be saturated prior to observing and/or measuring the signal intensity(ies). In the case of a single labeled subject nucleic acid, nonsaturation can be effected in a number of ways, for example, by stopping the hybridization, by providing insufficient subject nucleic acid, and/or by providing a sufficient amount of unlabeled nucleic acid which is sufficiently complementary to the reference chromosomes to competitively prevent saturation of sites therein by the labeled subject nucleic acid.

When there are two or more labeled subject nucleic acids, those subject nucleic acids can be hybridized *in situ* to the reference genome sequentially or simultaneously. Simultaneous *in situ* hybridization is preferred in that saturation of the targeted binding sites in the reference genome will not interfere with the procedure. When sequential *in situ* hybridization is used, it must be performed under conditions wherein the individual hybridizations are stopped well before the binding sites on the reference chromosomes are saturated.

Objects of this invention are to detect sequence copy number imbalances throughout an entire genome in one hybridization, to map gains and/or losses of sequences in a genome, and/or to provide a copy number karyotype of a subject genome.

Further, an object of this invention is to enable the detection of relative copy number differences that are common to a number of different cells and/or cell populations. For example, CGH methods can be used wherein DNAs extracted from cells of many different tumors are combined and labeled; the hybridization of those combined labeled DNAs to normal condensed chromosomes, provides for the rapid identification of only those copy number changes that occurred in most of the tumors. Less frequently occurring variations would be averaged out. Thus, this invention further provides for a CGH method wherein two or more of the subject nucleic acids that were extracted from different cells and/or from numbers of cells from different cell populations, are labeled the same, and hybridized to a reference spread under conditions wherein repetitive sequences are removed and/or suppressed and wherein sequence copy number differences that are common in said combined labeled nucleic acid sequences are determined.

Another object of this invention is to provide the means of cytogenetically analysing archived chromosomal material, that is, fixed material from, for example, biopsied tissue specimens, preferably cataloged and keyed to medical records of patients from whom the specimens were taken, and archaeological chromosomal material. Such chromosomal

material cannot, of course, be karyotyped according to traditional means in that no live cells are present to culture and from which to prepare chromosomal spreads. However, the nucleic acid can be extracted therefrom and amplified by a polymerase chain reaction (PCR) procedure or by a non-PCR procedure and tested by the methods of this invention.

This invention further provides for a method to detect simultaneously an ensemble of amplifications and/or deletions in a tumor wherein the results can be used to determine the subsequent behavior of that tumor. Said determination is made by associating the patterns of amplifications and/or deletions in tumor cells with the behavior of that tumor. Such associations can be made by testing, for example, as indicated immediately above, DNA from archived tumor tissue keyed to medical records, or when fresh tumor specimens are tested by CGH and the patients are followed. Further, such associations can be made with CGH methods wherein there are more than one subject cell and/or cell population, for example, one or more tumors.

Another object of this invention is to provide a method of analyzing cells from a suspected lesion at an early stage of development. An advantage of the methods of this invention is that only a few cells are necessary for the analysis. The early detection of amplifications and/or deletions in cells from a lesion allow for early therapeutic intervention that can be tailored to the extent of, for example, invasiveness known to be associated with such genetic rearrangements. Further, such early detection provides a means to associate the progression of the cells with the genetic rearrangements therein detected by the methods of this invention.

Tumors can be karyotypically heterogeneous containing therein various populations of cells each having different types of genetic rearrangements. As indicated above tumor cells are difficult to culture, and it is not clear that cultured cells are representative of the original tumor cell population. This invention provides the means to by-pass the culturing obstacle and allows genetic characterization of

tumor cells and thus, of the heterogeneity of tumors by testing cells from different subregions thereof according to the methods of this invention. Bulk extraction of the nucleic acid from many cells of a tumor can also be used to test for consistent amplifications and/or deletions within a tumor.

It is another object of this invention to provide methods of detecting amplifications and/or deletions of nucleic acid sequences wherein certain cell lines termed herein "antenna cell lines", are used to enhance the sensitivity of the detection.

It is still further an object of this invention to provide methods of prenatal or perinatal analysis wherein the nucleic acid of the child's cells is extracted and tested according to the methods of this invention. In one embodiment of CGH, such material is human and hybridized to a normal human metaphase spread to detect whether any deletions and/or amplifications are therein present, for example, an extra copy of chromosome 21, diagnostic for Down syndrome. Test kits for performing CGH methods are also provided.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 schematically illustrates the general approach used in performing the methods of this invention--Comparative Genomic Hybridization (CGH). The reference chromosome spread is hybridized with various nucleic acid mixtures, either simultaneously or at different times, to obtain the desired information. Representative mixtures could include unlabeled sequences designed to block sequences in the various other nucleic acid pools, for example, the high-copy repetitive sequences in human genomic DNA; unlabeled competitor nucleic acid to prevent saturation of the target sites for the labeled mixtures, for example, human genomic DNA within a factor of 10 of the concentration used for the labeled subject nucleic acids (see Figure 4); and one or more pools of sequences of different origin that are differently labeled so that their binding can be independently assessed, for example, tumor and normal genomic DNA (see Figures 5 and 6). The information on the sequence frequency of the labeled

pools is obtained by analysis of the intensity of the individual signals and/or the differences in ratios of intensities among the signals as a function of position along the reference chromosomes.

Figure 2 outlines general aspects of the CGH procedure used in Example 1, infra. The reference chromosome spread, in this example normal human chromosomes, is first hybridized for about one hour with a high concentration of unlabeled human genomic DNA (Figure 2A). That prehybridization blocks many of the high copy repetitive sequences in the chromosomes so that the high copy repetitive sequences in the labeled subject nucleic acid, in this case labeled tumor DNA, will not substantially contribute to the signal during the subsequent hybridization. The labeled tumor DNA, and perhaps some competitor DNA or other comparison nucleic acid are then hybridized to the target reference spread (Figure 2B). Cot-1 DNA can be included in the hybridization as in Example 1, below to block more effectively the centromeric repetitive sequences in the labeled subject nucleic acids.

Figure 2 is representative of one way of reducing signals from repetitive sequences. Other methods are detailed herein infra. In each of the CGH methods including the procedures outlined in the rest of the figures, some means of reducing the signal from the repetitive sequences is used, but not specifically indicated in the figures. It is important for CGH that the signal from each subject nucleic acid be dominated by sequences that bind to well-defined loci. Total suppression of the signal from the genomic repeats is not necessary, but the poorer the suppression, the less able the procedure is to detect small differences in sequence frequency.

Figure 3 further illustrates the procedure used in Example 1. As shown in Figure 3A, labeled human tumor DNA is hybridized to a normal human chromosome spread. [As indicated in the description for Figure 2, provisions were made to suppress the signal from the repetitive sequences although the provisions are not specifically indicated in the figure.

Example 1 details a preferred method to suppress the hybridization signals from repetitive sequences.] In this representative example, the tumor DNA is assumed to contain a region wherein some sequences are highly amplified, for example, an amplicon containing an oncogene. The amplified sequences in the tumor DNA may be clustered and integrated in some tumor chromosomes; they may be integrated into multiple places in the tumor genome; or, they may exist as extra-chromosomal elements. The sequences of the amplicon will map to some chromosomal location in the reference genome, which in this case is a normal human genome.

Figure 3B illustrates the kinetics of the build-up of the signal on a target reference chromosome. The signal builds more rapidly in the amplified region since more copies of those sequences are available for hybridization. If the reaction is stopped before the target chromosome is saturated, or if insufficient labeled DNA is added to achieve saturation, then the genomic region that was amplified in the tumor will appear higher in intensity on the normal chromosome as illustrated by the band with the denser shading on the left reference chromosome. The more intensely labeled region (band with the denser shading) indicates the location and extent of the amplicon as reflected in the reference genome. Thus, the amplification is detected without prior knowledge of its existence, and the origin of the amplified sequences is mapped in the normal human genome.

If the reaction illustrated in Figure 3 is allowed to proceed to saturation of the target sites, contrast is lost, as shown by the representative reference chromosome on the right wherein the amplicon cannot be distinguished. Thus, in this embodiment of CGH, it is important to stop the hybridization before saturation of the target or provide insufficient probe for saturation. The graphs schematically show the build-up of the hybridization signal in the region that was amplified (graph on right) and in the remainder that was unamplified (graph on left). The arrows connect the chromosomal regions with the times of observation on the kinetic curve.

Figure 4 illustrates an embodiment of CGH that avoids the potential saturation of the target as shown in the right portion of Figure 3B. In this representative example, the reference nucleic acid is a human chromosome spread; the subject nucleic acid is labeled tumor DNA (4A). If unlabeled human genomic DNA is included with the labeled tumor DNA in excess, in this case at a five-fold higher concentration than that of the labeled tumor DNA, then any saturation of the target will be due to a combination of labeled and unlabeled copies of the nucleic acid sequences, rather than just labeled copies as shown in the right portion of Figure 3B. [Once again, as indicated in Figures 2 and 3 the means of reducing the signal from repetitive sequences is not indicated in this figure, but it is assumed that a protocol is performed to remove substantially the repetitive sequences that would bind to multiple loci in the reference genome and/or to block such sequences from binding to the target.]

At the early stages of the reaction, the amplified region will build up faster than elsewhere in the chromosome (for example if the sequence is amplified five-fold, it would build up 5 times as fast) and will be detectable as in the left portion of Figure 3B. However as the reaction proceeds to saturation, the unamplified regions of the chromosome reach only one-fifth (1/5) of the intensity shown in the right portion of Figure 3B, because most of the sites are filled by unlabeled copies of the sequences. On the other hand, a sequence that was amplified five-fold in the tumor would reach one-half (1/2) of the saturation intensity since an equal number of labeled and unlabeled copies of those sequences are present. Thus, contrast is maintained according to this embodiment at all stages of the reaction (as shown in Figure 4B), although it changes as the reaction proceeds.

Figure 5 illustrates an embodiment of CGH designed to enhance its sensitivity in detecting small changes in copy number of various sequences. When a CGH procedure as indicated in Figure 4 is followed, intrinsic variation in the saturation levels, or rate of signal build-up at different positions in the reference genome may not be indicative of

abnormal gain or loss of sequences. Such intrinsic variations would interfere with interpretation of intensity differences as indicating differences in copy number of the sequences. This CGH embodiment overcomes that potential problem by providing a mixture of labeled subject nucleic acid, in this case tumor DNA labeled with a green fluorochrome, and a differently labeled competitor nucleic acid in this case normal human genomic DNA labeled with a red fluorochrome. The two differently labeled DNAs are simultaneously hybridized to the chromosome spread. [Once again, removal of the repetitive sequences and/or blocking of the signal therefrom is performed but not illustrated.] Changes in the ratio of green to red along each of the chromosomes in the reference spread then indicate regions of increased or decreased sequence copy number in the tumor. Those ratio changes may result in color variations from red to yellow to green on the reference spread.

Figure 6 graphically and schematically explains the kinetics underlying the CGH embodiment illustrated in Figure 5. In the center is one of the chromosomes of the reference chromosome spread, a normal human chromosome in this case. The darkness of the shading on the reference chromosome shows the ratio of green to red intensity along the chromosome.

In the amplified region, the green/red ratio is much higher than in the normal region, whereas in the deleted region the green/red ratio is less than in the normal region. The arrows from examples of each of the different green/red intensity regions point to kinetic curves that indicate the build-up of green (solid line for the tumor DNA) and red (dashed line for the normal DNA) signals during the hybridization. In the normal region, upper left graph, the red and green signals build together. (They have been normalized to be equal for the purposes of this explanation.) In the amplified region, upper right, the green (tumor) signal builds up much more rapidly than the red (normal) signal, the green/red ratio being approximately the level of amplification (given the normalization to the normal part of the chromosome).

In the lower left of Figure 6, the signal build-up for the duplicated region is shown; the green (tumor) signal is 50% brighter than the red (normal) signal. In the lower right, the build-up for a deleted region is schematically described; the green (tumor) signal is 50% dimmer than the red (normal) signal. The ratio approach of this CGH embodiment further normalizes for the frequent finding that hybridization to some chromosomes in a spread is intrinsically brighter than that for others because of differences in the local hybridization environment.

Figure 7 graphically illustrates the correlation of the number of X chromosomes in five fibroblast cell lines and the average green-to-red ratio of the X chromosome(s) relative to the same ratio for the autosomes.

Figure 8 illustrates green-to-red fluorescence ratio profiles of chromosomes 1, 9, 11, 16 and 17 after comparative genomic hybridization with breast cancer cell line 600PE (green) and with a normal DNA (red). The profiles reflect the relative copy number of the chromosomal regions. Fluorescence *in situ* hybridization (FISH) with 16p and 16q cosmid probes to interphase and metaphase 600PE cells indicated that there were two signals with 16p cosmid probes and one signal from the 16q cosmid probes. That information on the absolute copy number of those loci provided by FISH permits interpretation of the ratio 1.0 as indicating that there are two copies of the sequence throughout the genome.

The dip in the profile at 1p34 through 1p36 may represent a previously unsuspected small interstitial deletion; however, that observation has not yet been independently verified with specific probes for that region.

Centromeric and heterochromatic regions of the genome are not included in the analysis because the Cot-1 DNA partially blocks signals in those regions, and the large copy number polymorphisms between individual sequences at those loci effect unreliable ratio data.

Figure 9A and 9B respectively provide green-to-red fluorescence ratio profiles of chromosome 8 (Figure 9A) and chromosome 2 (Figure 9B) after comparative genomic

hybridization respectively with COLO 320 HSR (human colon adenocarcinoma cell line) and NCI H69 (small cell lung carcinoma cell line) cell line DNAs (green) and with normal human DNA (red).

In Figure 9A, the myc locus at 8q24 shows a highly elevated green-to-red ratio, which is consistent with the known high level amplification of myc in the COLO 320HSR cell line.

In Figure 9B, three regions of amplification are seen on chromosome 2. The signal at 2p24 corresponds to the location of N-myc known to be amplified in the NCI-H69 cell line. The two other regions with a highly increased green-to-red fluorescence ratio, at 2p21 and 2q21, were not previously known to be amplified in the NCI-H69 cell line.

DETAILED DESCRIPTION

Comparative Genomic Hybridization (CGH) has also been termed Copy Ratio Reverse Cytogenetics (CRRC), competition hybridization and quantitative in situ ratio karyotyping (QUIRK). Further, in the embodiment wherein fluorochromes are used as labels, it has been termed competition FISH (fluorescence in situ hybridization). CGH specifically provides methods whereby amplifications, duplications and/or deletions can be identified in an immediate overview of a genome.

CGH provides methods for determining variations in the copy number of different elements in a mixture of nucleic acid sequences (for example, genomic DNA isolated from a tumor) as a function of the location of those sequences in the genome of a reference organism (for example, the genome of a normal cell from the same species). The methods comprise the use of in situ hybridization of the nucleic acid sequence mixture to a chromosome spread of the reference organism, and measuring the intensity of the hybridization at different locations along the target chromosomes. Exemplary methods are schematically outlined in Figures 1-6. Those illustrative examples are not exhaustive but suggest the wide range of variations and other uses of the basic approach.

As the figure descriptions indicate, it is critical that signals from repetitive sequences do not dominate the signal from the subject nucleic acid pool, and that they be removed from the pool or that their signals be suppressed as necessary. It is preferred to exclude sequences from the hybridization or block sequences in the hybridization mixture that could bind to multiple clearly separated positions on the chromosomes, for example, sites that are on different chromosomes, or that are on the same chromosome but are well-separated. In many applications of CGH, it is the high copy repetitive sequences, such as Alu, Kpn, Lines, and alpha-satellites among others, that are removed from the labeled subject nucleic acid and/or which are blocked and/or the binding sites therefor are blocked. Described herein are methods to remove and/or block those repetitive signals. It should be noted that nucleic acid sequences in the labeled nucleic acid that bind to single copy loci are substantially retained in the hybridization mixture of labeled subject nucleic acids, and such single copy sequences as well as their binding sites in the reference chromosome spread remain substantially unblocked relative to the repetitive sequences that bind to multiple loci (that is, loci that are visually distinguishable) both before and during the hybridization.

The methods of this invention provide the means to identify previously unknown regions of amplification and deletion. For example, one embodiment of CGH as detailed in Example 1 herein provides an efficient method that gives an immediate overview of a genome identifying all regions that are amplified greater than about five-fold to ten-fold as well as at least large deletions. More sensitive embodiments that can identify smaller amplifications and deletions are also disclosed.

Nanogram quantities of the subject nucleic acids are required for the CGH methods of this invention. Paraffin embedded tumor sections can be used as well as fresh or frozen material. Snap frozen material from normal and malignant tissue are preferred for mRNA isolation.

Standard procedures can be used to isolate the required nucleic acid from the subject cells. However, if the nucleic acid, for example, DNA or mRNA, is to be extracted from a low number of cells (as from a particular tumor subregion) or from a single cell, it is necesary to amplify that nucleic acid, by a polymerase chain reaction (PCR) procedure or by a non-polymerase chain reaction (non-PCR) procedure. PCR and preferred PCR procedures are described infra. Exemplary non-PCR procedures include the ligase chain reaction (LCR) and linear amplification by use of appropriate primers and their extension (random priming).

Some of the various embodiments of CGH are illustrated, particularly in Figures 1-6. In the embodiment illustrated in Figures 5 and 6, wherein a subject nucleic acid, in this case, human genomic DNA, that is labeled differently from another subject nucleic acid, amplifications and/or deletions are indicated by a change in ratio between the different signals, rather than just a change in signal intensity.

The representative examples concerning CGH of Examples 1, 2 and 3 below involve the hybridizations of tumor cell line DNA to normal human metaphase spreads. However, there are many permutations and combinations of pairwise and multiple hybridizations of different nucleic acids from different genomes all of which are considered to be within the scope of this invention.

For example, CGH could be used to hybridize labeled DNA from a tumor cell line to metaphase spreads of that same cell line to estimate the level and pattern of amplification in each cell line, comparing those results to hybridizations of said tumor cell line DNA to a normal human metaphase spread. Alternatively, labeled tumor cell line DNA and differently labeled human genomic DNA could be simultaneously hybridized to a metaphase spread of a tumor cell line metaphase spread. Further, DNA from a primary tumor and that from its metastasis could be differently labeled and hybridized in a CGH method to a normal human metaphase or to a

related tumor cell line metaphase. Those are just some of the many examples of CGH.

Although the examples herein concern the hybridizations of the DNA from breast cancer cell lines and primary tumors to normal human metaphase spreads, it will be clear to anyone skilled in the art that CGH is not limited to studying genomes of cancer cells or to the results of hybridizing abnormal genomes to normal genomes. CGH permits the comparison of nucleic acid sequence copy frequencies of any two or more genomes, even genomes of different species if their nucleic acid sequences are sufficiently complementary to allow for meaningful interpretation. It should be noted regarding interspecies comparisons that the information obtained by CGH includes not only an assessment of relative copy number but also that of sequence divergence.

It will also be clear to those skilled in the art that hybridization with nucleic acid other than chromosomal DNA, such as messenger RNA (mRNA) or complementary DNA (cDNA) of subject cells can be used to determine the location and level of expression of genes in those cells. Conventional methodology is used to extract mRNA from a cell or cell population, and to synthesize in vitro cDNA by reverse transcription.

CGH does not require the preparation of condensed chromosomes, for example, metaphase, prophase or other condensed chromosomal states, of the subject genomes. Thus, genomes from which metaphase, prophase or otherwise condensed chromosomal spreads are difficult, time-consuming or not possible to prepare at least in good quality, for example, genomes of tumor cells or fetal cells can be studied by CGH.

In CGH, labeled subject nucleic acids, for example, labeled tumor DNA, is hybridized to a reference genome, for example, a normal human metaphase spread, under conditions in which the signal from amplified, duplicated and/or deleted nucleic acid sequences from the labeled nucleic acid can be visualized with good contrast. Such visualization is accomplished by suppressing the hybridization of repetitive sequences that bind to multiple loci including the high copy

interspersed and clustered repetitive sequences, such as, Alu, Kpn, Lines, alpha-satellites among others, using unlabeled total human genomic nucleic acid, preferably DNA, and/or the repeat-enriched (Cot-1) fraction of genomic DNA, and/or by removing such repetitive sequences from the hybridization mixture. In providing the detection sensitivity required, the extent of suppression of the hybridization of repetitive sequences and/or removal thereof can be adjusted to the extent necessary to provide adequate contrast to detect the differences in copy number being sought; for example, subtler copy number changes may require the suppression or removal of lower level repetitive sequences.

When combining more than one labeled nucleic acid in a hybridization mixture, the relative concentrations and/or labeling densities may be adjusted for various purposes. For example, when using visual observation or photography of the results, the individual color intensities need to be adjusted for optimum observability of changes in their relative intensities. Adjustments can also be made by selecting appropriate detection reagents (avidin, antibodies and the like), or by the design of the microscope filters among other parameters. When using quantitative image analysis, mathematical normalization can be used to compensate for general differences in the staining intensities of different colors.

The kinetics of the CGH hybridizations are complicated. Since the subject nucleic acids are frequently double stranded, complementary sequences will reassociate in the hybridization mix as well as hybridizing to the target. Such reassociation may result in a more rapid decrease in concentration of the high copy sequences than the low copy ones, thereby making the signal intensity variations on the reference chromosomes less pronounced than the copy differences in the original subject DNAs. In addition, non-specific binding of the labeled subject DNAs to the slide, coverslip, etc. may generally reduce the concentration of that labeled subject nucleic acid during the hybridization. Those skilled in the art will recognize numerous methods of

optimizing the quantitative aspects of CGH, such as, mathematical correction of digital images, supplying freshly denatured subject DNA during the hybridization, and adding unlabeled genomic DNA in excess to dominate the reassociation rates. The term "saturation" is defined in the context of hybridization kinetics.

The resolution of CGH is presently at a level that can be seen through a light microscope, as is traditional cytogenetic staining. Thus, if a small sequence in a subject nucleic acid is amplified, to be seen as a signal in a subject genome, it must be amplified enough times for its signal to be able to be visualized under a light microscope. For example, the locus for erbB-2 which is relatively small (very approximately, a few hundred kb), needs to be amplified at least greater than five times to be visually distinguishable under a light microscope when the CGH embodiment used in Example 1 is employed. On the other hand, if a large section of a chromosome is present at increased frequency in a subject nucleic acid, the signal from that region would show up in the reference genome at a much lower level of amplification.

The term "labeled" is herein used to indicate that there is some method to visualize nucleic acid fragments that are bound to the target, whether or not the fragments directly carry some modified constituent. A section <u>infra</u> entitled "Labeling the Nucleic Acid Fragments of the Subject Nucleic Acids" describes various means of directly labeling the probe and other labeling means by which the bound probe can be detected.

The phrase "antenna cell line" is herein used to indicate a reference genome that has one or more known significant genetic aberrations, for example, a cell line known to have an oncogene that is highly amplified, for example, in large homogeneously staining regions (HSRs). The amplified regions of that cell line would thus provide a much bigger target site than a normal chromosome spread. Thus, observation of the signal from such a large target site would be easier in that on average the signal would be brighter from amplified target sequences in the reference genome as provided

by such an antenna cell line. A subject nucleic acid extracted from, for example, a number of tumor cells, could be tested by a CGH hybridization to such an antenna cell line to see if it also contained amplification(s) of the oncogene known to be amplified in the cell line.

When an antenna cell line is used as the reference genome, there are instances wherein it can be used in interphase rather than as a chromosome spread. For example, if one is checking to see if a certain oncogene is amplified or not in the subject nucleic acid, interphase CGH is sufficient. However, the maximum amount of information is provided when condensed chromosome spreads are used.

A base sequence at any point in the genome can be classified as either "single-copy" or "repetitive". For practical purposes the sequence needs to be long enough so that a complementary probe sequence can form a stable hybrid with the target sequence under the hybridization conditions being used. Such a length is typically in the range of several tens to hundreds of nucleotides.

A "single-copy sequence" is that wherein only one copy of the target nucleic acid sequence is present in the haploid genome. "Single-copy sequences" are also known in the art as "unique sequences". A probe complementary to a single-copy sequence has one binding site in haploid genome. A "repetitive sequence" is that wherein there is more than one copy of the same target nucleic acid sequence in the genome. Each copy of a repetitive sequence need not be identical to all the others. The important feature is that the sequence be sufficiently similar to the other members of the family of repetitive sequences such that under the hybridization conditions being used, the same fragment of probe nucleic acid is capable of forming stable hybrids with each copy.

Herein, the terms repetitive sequences, repeated sequences and repeats are used interchangeably.

The phrase "metaphase chromosomes" in herein defined to encompass the concept of "condensed chromosomes" and is defined to mean not only chromosomes condensed in the prophase or metaphase stage of mitosis but any condensed chromosomes,

for example, those condensed by premature chromosome condensation or at any stage in the cell cycle wherein the chromosome can be visualized as an individual entity. It is preferred that the chromosomes in the reference genome be as long as possible but condensed sufficiently to be visualized individually.

A subject nucleic acid is herein considered to be the same as another nucleic acid if it is from a member of the same sex of the same species and has no significant cytogenetic differences from the other nucleic acid. For example, the DNA extracted from normal lymphocytes of a human female is considered for the purposes of this invention to be the same nucleic acid as that of DNA from normal cells of a human female placenta.

The following abbreviations are used herein:

<u>Abbreviations</u>

AAF	-	N-acetoxy-N-2-acetyl-aminofluorene
ATCC	-	American Type Culture Collection
BN	-	bicarbonate buffer with NP-40
Brd/ Urd	-	bromodeoxyuridine
BRL	-	Bethesda Research Laboratories
bp	-	base pair
CCD	-	charge coupled device
CGH	-	Comparative Genomic Hybridization
Chr.	-	chromosomal
CML	-	chronic myelogenous leukemia
CRRC	-	Copy Ratio Reverse Cytogenetics
DAPI	-	4,6-diamidino-2-phenylindole
dATP	-	deoxyadenosine triphosphate
DCS	-	as in fluorescein-avidin DCS (a commercially available cell sorter grade of fluorescein Avidin D)
dCTP	-	deoxycytosine triphosphate

dGTP	-	deoxyguanosine triphosphate
DI	-	DNA index
DM	-	double minute chromosome
dNTP	-	deoxynucleotide triphosphate
dTTP	-	deoxythymidine triphosphate
dUTP	-	deoxyuridine triphosphate
EDTA	-	ethylenediaminetetraacetate
E/P	-	estrogen/progesterone
FISH	-	fluorescence _in situ_ hybridization
FACS	-	fluorescence-activated cell sorting
FITC	-	fluorescein isothiocyanate
HPLC	-	high performance liquid chromatography
HSR	-	homogeneously staining region
ISCN	-	International System for Cytogenetic Nomenclature
IB	-	isolation buffer
kb	-	kilobase
kDa	-	kilodalton
LCR	-	ligase chain reaction
LOH	-	loss of heterozygosity
Mb	-	megabase
met.	-	metastasis
min	-	minute
ml	-	milliliter
mM	-	milliMole
mm	-	millimeter
ng	-	nanogram
NIGMS	-	National Institute of General Medical Sciences
NP-40	-	non-ionic detergent commercially available from Sigma as Nonidet P-40 (St. Louis, MO)

PBS	-	phosphate-buffered saline
PCR	-	polymerase chain reaction
PHA	-	phytohemagglutinin
PI	-	propidium iodide
Pl.	-	pleural
PMSF	-	phenylmethylsulfonyl fluoride
PN buffer	-	mixture of 0.1 M NaH_2PO_4 and 0.1 M Na_2HPO_4, pH 8; 0.1% NP-40
PNM buffer	-	Pn buffer plus 5% nonfat dry milk (centrifuged); 0.02% Na azide
QUIPS	-	quantitative image processing system
QUIRK	-	quantitative <u>in situ</u> ratio karyotyping
Rb-1	-	retinoblastoma tumor suppressor gene
RFLP	-	restriction fragment length polymorphism
RPM	-	revolutions per minute
SD	-	Standard Deviation
SDS	-	sodium dodecyl sulfate
SSC	-	0.15 M NaCl/0.015 M Na citrate, pH 7
Td	-	doubling time
ug	-	microgram
ul	-	microliter
um	-	micrometer
uM	-	micromole
VNTR	-	variable number tandem repeat

Resolution of differences in copy number can be improved by the use of image analysis and by averaging the results from hybridizations of a subject nucleic acid to multiple condensed chromosome spreads. Using such methods, the background signal (noise) can be differentiated from actual nucleic acid sequence copy number differences.

Image Analysis:

An image analysis system, preferably computer-assisted, can be used to enhance and/or accurately quantitate the intensity differences between and/or among the signals from a hybridization and the background staining differences for more accurate and easier interpretation of results. Image analysis and methods to measure intensity are described, for example, in Hiraoka et al., Science, 238: 36-41 (1987) and Aikens et al., Meth. Cell Biol., 29: 291-313 (1989). In such an image analysis system, it is preferred to use a high quality CCD camera whose intensity response is known to be linear over a wide range of intensities.

The components of a particular quantitative image processing system (QUIPS) are described in Example 1 under the subheading Fluorescence Microscopy and Interpretation of Results. As exemplified in Example 1, a computer-assisted image analysis system with a filterwheel is used so that the images from the signals and counterstaining of the DNA are superimposed on one image. Pseudocolors, that is, colors that are not exactly spectrally converted, can be displayed. Contrast stretching, wherein the differences between the intensity levels of the signals and background staining differences are enhanced by adjusting controls of the image analysis system. Thresholding can also be used wherein the background staining can be assigned a value close to zero so it would barely appear in the processed image from such a system. Similarly, computer analysis permits substraction of background, smoothing of fluctuations in the signals, accurate intensity and ratio calculations and the ability to average signals on chromosomes in multiple spreads.

Absolute Copy Numbers:

Hybridization of the subject DNAs to the reference chromosomes gives information on relative copy numbers of sequences. Some additional normalization is required to obtain absolute copy number information. One convenient method to do this is to hybridize a probe, for example a cosmid specific to some single locus in the normal haploid

genome, to the interphase nuclei of the subject cell or cell population(s) (or those of an equivalent cell or representative cells therefrom, respectively). Counting the hybridization signals in a representative population of such nuclei gives the absolute sequence copy number at that location. Given that information at one locus, the intensity (ratio) information from the hybridization of the subject DNA(s) to the reference condensed chromosomes gives the absolute copy number over the rest of the genome. In practice, use of more than one reference locus may be desirable. In this case, the best fit of the intensity (ratio) data through the reference loci would give a more accurate determination of absolute sequence copy number over the rest of the genome.

Thus, the CGH methods of this invention combined with other well-known methods in the art can provide information on the absolute copy numbers of substantially all RNA or DNA sequences in subject cell(s) or cell population(s) as a function of the location of those sequences in a reference genome. For example, one or more chromosome-specific repeat sequence or high complexity painting probes can be hybridized independently to the interphase nuclei of cells representative of the genomic constitution of the subject cell(s) or cell population(s). Whole chromosome painting probes are now available for all the human chromosomes [Collins et al., Genomics, 11: 997-1006 (1991)]. Specific repeat-sequence probes are also available [Trask et al., Hum. Genet., 78: 251 (1988) and references cited therein; and commercially available from Oncor (Gaithersburg, MD, USA)]. Hybridization with one or more of such probes indicates the absolute copy numbers of the sequences to which the probes bind.

For such interphase analysis, painting probes with a complexity of from about 35 kb to about 200 kb, are preferred; probes from about 35 kb to about 100 kb are further preferred; and still more preferred are probes having a complexity of from about 35 kb to 40 kb, for example, a cosmid probe. Exemplary of such locus-specific painting probes are any

cosmid, yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), and/or p1 phage probes as appropriate, preferably to the arms of a selected chromosome. Such cosmid probes, for example, are commercially available from Clontech [South San Francisco, CA (USA)] which supplies cosmid libraries for all the human chromosomes. Another example of a cosmid probe that could be used in such methods of this invention would be a 3p cosmid probe called cC13-787 obtained from Yusuke Nakamura, M.D., Ph.D. [Division of Biochemistry, Cancer Institute, Toshima, Tokyo, 170, Japan]. Its isolation and mapping to 3p21.2-p21.1 is described in Yamakawa et al., Genomics, 9(3): 536-543 (1991). Another example would be a 3q cosmid probe named J14R1A12 obtained from Wen-Lin Kuo [Biomedical Department, P.O. Box 5507 (L-452), Lawrence Livermore National Laboratory Livermore, CA 94550 (USA)]. For interphase analysis, preferred repeat sequence probes are centromeric-specific and/or peri-centromeric-specific repeat sequence probes. Such a centromeric-probe is, for example, the chromosome 17 peri-centromeric repeat probe (cosmid ck17.10) and the alpha satellite repeat probe for the centromeric region of chromosome 8, both of which are described in Example 1 *infra*. A variety of repeat sequence probes are commercially available from Oncor [Gaithersburg, MD (USA)]. However, the locus-specific painting probes are preferred over the repeat sequence probes for the methods of this invention to determine absolute copy numbers of nucleic acid sequences.

Further, when the subject nucleic acid sequences are DNA, the reference copy numbers can be determined by Southern analysis. When the subject nucleic acid sequences are RNA, the reference copy numbers can be determined by Northern analysis.

Those reference copy numbers or reference frequencies provide a standard by which substantially all the RNA or DNA sequences in the subject cell(s) or cell population(s) can be determined. CGH methods are used to determine the relative copy numbers of the rest of the sequences. However, absolute copy numbers require a standard

against which the results of CGH can be determined. Otherwise the CGH procedures would have to be highly standardized and quantitated to see differences in the absolute copy numbers of sequences in a genome, for example, haploidy, triploidy, octaploidy, wherein there are 1, 3 and 8 copies of each of the chromosomes, respectively.

PCR and Microdissection:

The mechanics of PCR are explained in Saiki et al., Science, 230: 1350 (1985) and U.S. Patent Nos. 4,683,195, 4,683,202 (both issued July 18, 1987) and 4,800,159 (issued January 24, 1989).] PCR offers a rapid, sensitive and versatile cell-free molecular cloning system in which only minute amounts of starting material are required.

A preferred PCR method to amplify the subject nucleic acids for testing by CGH is a PCR adapter-linker amplification [Saunders et al., Nuc. Acids Res., 17 9027 (1990); Johnson, Genomics, 6: 243 (1990) and PCT 90/00434 (published August 9, 1990)]. The labeled subject nucleic acid could be produced by such a adapter-linker PCR method from a few hundred cells; for example, wherein the subject nucleic acid is tumor DNA, the source DNA could be a few hundred tumor cells. Such a method could provide a means to analyse by CGH clonal sub-populations in a tumor.

Another further preferred PCR method is a method employing a mixture of primers described in Meltzer et al., "Rapid Generation of Region Specific Probes by Chromosome Microdissection and their Application: A Novel Approach to Identify Cryptic Chromosomal Rearrangements," Nature--Genetics, 1(1): 24-28 (April 1992). Microdissection of sites in the reference metaphase spread that produce signals of interest in CGH, would permit PCR amplification of nucleic acid sequences bound at such sites. The amplified nucleic acid could then be easily recovered and used to probe available libraries, as for example, cosmid libraries, so that the amplified sequences could be more rapidly identified.

High copy repetitive sequences can be suppressed in amplifying the subject nucleic acid by PCR. The PCR primers

used for such a procedure are complementary to the ends of the repetitive sequences. Thus, upon proper orientation, amplification of the sequences flanked by the repeats occurs. One can further suppress production of repetitive sequences in such a PCR procedure by first hybridizing complementary sequences to said repetitive sequences wherein said complementary sequences have extended non-complementary flanking ends or are terminated in nucleotides which do not permit extension by the polymerase. The non-complementary ends of the blocking sequences prevent the blocking sequences from acting as a PCR primer during the PCR process. Primers directed against the Alu and L1 repetitive DNA families have allowed the selective amplification of human sequences by interspersed repetitive sequence PCR (IRS-PCR) [Nelson et al., PNAS, 86: 6686 (1989); Ledbetter et al., Genomics, 6: 475 (1990)].

Archived Material

An important aspect of this invention is that nucleic acids from archived tissue specimens, for example, paraffin-embedded or formalin-fixed pathology specimens, can be tested by the methods of CGH. Said nucleic acid cannot, of course, be prepared into chromosome spreads for traditional cytogenetic chemical staining. Also, it is difficult for large enough restriction fragments to be extracted from such material for other conventional research tools, such as Southern analysis. However, the nucleic acid from such specimens can be extracted by known techniques such as those described in Greer et al., Anatomic Pathology, 95(2): 117-124 (1991) and Dubeau et al., Cancer Res., 46: 2964-2969 (1986), and if necessary, amplified for testing by various CGH methods. Such nucleic acid can be amplified by using a polymerase chain reaction (PCR) procedure (described above), for example, by the method described in Greer et al., supra wherein DNA from paraffin-embedded tissues is amplified by PCR.

A particular value of testing such archived nucleic acid is that such specimens are usually keyed to the medical

records of the patients from whom the specimens were taken. Therefore, valuable diagnostic/prognostic associations can be made between the revealed cytogenetic state of patients' nucleic acid material and the medical histories of treatment and outcome for those patients. For example, information gathered by CGH can be used to predict the invasiveness of a tumor based upon its amplification and/or deletion pattern matched to associations made with similar patterns of patients whose outcomes are known.

Analogously, other nucleic acid that is fixed by some method, as, for example, archaelogical material preserved through natural fixation processes, can also be studied by CGH procedures. As indicated above, copy number differences between species provide information on the degree of similarity and divergence of the species studied. Evolutionarily important linkages and disjunctions between and among species, extant or extinct, can be made by using the methods of CGH.

Tumor Cytogenetics

CGH provides the means to assess the association between gene amplification and/or deletion and the extent of tumor evolution. Correlation between amplification and/or deletion and stage or grade of a cancer may be prognostically important because such information may contribute to the definition of a genetically based tumor grade that would better predict the future course of disease with more advanced tumors having the worst prognosis. In addition, information about early amplification and/or deletion events may be useful in associating those events as predictors of subsequent disease progression. Gene amplification and deletions as defined by CGH to, for example, normal metaphase spreads (genomic site, intensity of the signal and/or differences in signal ratios, and number of different genomic sites at which the copy number differences occur) can be associated with other known parameters such as tumor grade, histology, Brd/Urd labeling index, hormonal status, nodal involvement, tumor size, survival duration and other tumor properties available

from epidemiological and biostatistical studies. For example, tumor DNA to be tested by CGH could include atypical hyperplasia, ductal carcinoma in situ, stage I-III cancer and metastatic lymph nodes in order to permit the identification of associations between amplifications and deletions and stage.

The associations made may make possible effective therapeutic intervention. For example, consistently amplified regions may contain an overexpressed gene, the product of which may be able to be attacked therapeutically (for example, the growth factor receptor tyrosine kinase, $p185^{HER2}$).

CGH hybridizations of nucleic acids from cells of primary cancers that have metastasized to other sites can be used to identify amplification and/or deletion events that are associated with drug resistance. For example, the subject nucleic acids to be analysed could be selected so that approximately half are from patients whose metastatic disease responded to chemotherapy and half from patients whose tumors did not respond. If gene amplification and/or deletion is a manifestation of karyotypic instability that allows rapid development of drug resistance, more amplification and/or deletion in primary tumors from chemoresistant patients than in tumors in chemosensitive patients would be expected. For example, if amplification of specific genes is responsible for the development of drug resistance, regions surrounding those genes would be expected to be amplified consistently in tumor cells from pleural effusions of chemoresistant patients but not in the primary tumors. Discovery of associations between gene amplification and/or deletion and the development of drug resistance may allow the identification of patients that will or will not benefit from adjuvant therapy.

Once a new region of amplification or deletion has been discovered by CGH, it can be studied in more detail using chromosome-specific painting [Pinkel et al., PNAS (USA), 85: 9138-9142 (1988); EP Publication No. 430,402 (June 5, 1991)] with a collection of probes that span the amplified or deleted region. Probes to amplified regions will show more signals than centromeric signals from the same chromosome, whereas

probes to nonamplified regions will show approximately the same number of test and centromeric signals. For example, the amplified regions on 17q22-23 and 20qter (discussed as newly discovered regions of amplification in Example 1) show variability in size from tumor to tumor using CGH (the 17q22-23 region more markedly); it can be expected that the region containing the important gene(s) can be narrowed by mapping the regions of amplification in multiple tumors in more detail to find the portion that is amplified in all cases. Probes for those studies can be selected, for example from specific cosmid libraries produced by the National Laboratory Gene Library Project and/or from the National Institute of Health (NIH) genomic research projects.

The c-erbB-2 oncogene, also referred to as HER-2 or neu, encodes for a 185 kilodalton (Kd) protein. Studies have reported c-erbB-2 gene amplification in human mammary tumor cell lines. [Kràus et al., EMBO J. 6: 605-610 (1987); van de Vijver et al., Mol. Cell Biol., 7: 2019-2023 (1987).] Also, c-erbB-2 gene amplification in human breast cancer has been shown to be associated with disease behavior, and may be a predictor of clinical outcome. [Slamon et al., Science, 235: 177-182 (1987); Berger et al., Cancer Res., 48: 1238-1243 (1988); Zhou et al., Cancer Res., 47:6123-6125 (1987); and Venter et al., Lancet, 11: 69-71 (1987)]. C-erbB-2 has also been shown to be amplified in ovarian cancers. [Alitalo and Schwab, Advances in Cancer Res., 47: 235-281 (1986).]

C-myc is a proto-oncogene which is the cellular homolog of the transforming gene of the chicken retrovirus MC29. In humans, c-myc lies on the long arm of chromosome 8, at band 124, and spans about 5 kilobase pairs. The myc protein is a phosphoprotein present in the nucleus. The normal function of c-myc is unknown; however, it also certainly plays a role in cell division, and is expressed in normally growing cells as well as in tumor cells. It is now widely believed that translocations involving c-myc lead to altered transcription of the gene, contributing to malignant transformation.

Sequences from N-myc member of the myc gene family have been shown to be amplified as much as a thousandfold in some neuroblastomas. N-myc amplifications are usually seen in the later stage III and IV tumors. Some small-cell lung carcinomas also have amplified myc genes in double minute chromosomes (DMs) and homogeneously staining regions (HSRs). Myc has also been shown to be amplified in colon cancer. [Alitalo and Schwab, supra.] Again such amplifications are found in late stages of tumor development, in the so-called variant cells that exhibit a more malignant behavior. Amplifications can involve either c-myc, N-myc or another member of the myc gene family, L-myc. [Watson et al., supra at pp. 1084-1086].

In addition, overexpression has been observed for the p-glycoprotein gene family associated with multi-drug resistance and for drug metabolizing enzymes such as P450 containing enzymes and glutathione S-transferase. [Fairchild and Cowan, J. Radiation Oncol. Biol. Phys., 20: 361-367 (1990).]

Identification of amplified and/or deleted genes is important to the management of cancer, for example, breast cancer, for several reasons:

1) to improve prognostication;

2) to detect amplification and/or deletion events that are associated with the development of drug resistance; and

3) to improve therapy.

For example, in regard to improving prognostication, in breast cancer the amplification of oncogenes, such as int-2, erbB-2 and myc occur frequently and have been associated with aggressive growth and poor prognosis in some studies. [Schwab and Amier, Genes, Chromosomes & Cancer, 1: 181-193 (1990).] In regard to reason (2), gene amplification has clearly been shown to lead to drug resistance in vitro (for example, amplification of the dihydrofolate reductase gene confers resistance to methotrexate), and is likely to occur in patients undergoing therapy as well (for example, as a result of over expression of glutathione S-transferase and p-

glycoprotein). [Fairchild and Cowan, supra]. Thus, the identification of resistance-linked genes would have a major impact on therapy by allowing therapy modification as resistance-related gene amplification occurs. Therapy could be improved by targeting for specific therapy, tumors that overexpress specific amplified genes.

Prenatal Diagnosis

Prenatal screening for disease-linked chromosome aberrations (e.g., trisomy 21) is enhanced by the rapid detection of such abberrations by the methods and compositions of this invention. CGH analysis is particularly significant for prenatal diagnosis in that it yields more rapid results than are available by cell culture methods.

Removal of Repetitive Sequences and/or Disabling the Hybridization Capacity of Repetitive Sequences

The following methods can be used to remove repetitive sequences and/or disable the hybridization capacity of such repetitive sequences. Such methods are representative and are expressed schematically in terms of procedures well known to those of ordinary skill the art, and which can be modified and extended according to parameters and procedures well known to those in the art.

Bulk Procedures. In many genomes, such as the human genome, a major portion of distributed (or shared) repetitive DNA is contained in a few families of highly repeated sequences such as Alu. These methods primarily exploit the fact that the hybridization rate of complementary nucleic acid strands increases as their concentration increases. Thus, if a mixture of nucleic acid fragments is denatured and incubated under conditions that permit hybridization, the sequences present at high concentration will become double-stranded more rapidly than the others. The double-stranded nucleic acid can then be removed and the remainder used in the hybridizations. Alternatively, the partially hybridized mixture can be used as the subject nucleic acid, the double-stranded sequences being unable to bind to the target. The following are methods

representative of bulk procedures that are useful for disabling the hybridization capacity of repetitive sequences or removing those sequences from a mixture.

Self-reassociation. Double-stranded nucleic acid in the hybridization mixture is denatured and then incubated under hybridization conditions for a time sufficient for the high-copy sequences in the mixture to become substantially double-stranded. The hybridization mixture is then applied to the reference chromosome spread. The remaining labeled single-stranded copies of the highly repeated sequences may bind throughout the reference chromosome spread producing a weak, widely distributed signal.

Use of blocking nucleic acid. Unlabeled nucleic acid sequences which are complementary to those sequences in the hybridization mixture whose hybridization capacity it is desired to inhibit are added to the hybridization mixture. The subject nucleic acids and blocking nucleic acid are denatured, if necessary, and incubated under appropriate hybridization conditions. The sequences to be blocked become double-stranded more rapidly than the others, and therefore are unable to bind to the reference spread when the hybridization mixture is applied to the spread. In some cases, the blocking reaction occurs so quickly that the incubation period can be very short, and adequate results can be obtained if the hybridization mix is applied to the spread immediately after denaturation. Further, the probe and the target can be simultaneously denatured in some cases. A blocking method is generally described in the context of Southern analysis by Sealy et al., "Removal of Repeat Sequences form Hybridization Probes", Nucleic Acid Research, 13:1905 (1985). Examples of blocking nucleic acids include genomic DNA, a high-copy fraction of genomic DNA and particular sequences as outlined below.

i. Genomic DNA. Genomic DNA contains all of the nucleic acid sequences of the organism in proportion to their copy-number in the genome. Thus, adding genomic DNA to the hybridization mixture increases the concentration of the

high-copy repeat sequences more than low-copy sequences, and therefore is more effective at blocking the former.

ii. High-copy fraction of genomic DNA. Fractionating the genomic DNA to obtain only the high-copy sequences and using them for blocking can be done, for example, with hydroxyapatite as described below.

Removal of Sequences.

Hydroxyapatite. Single- and double-stranded nucleic acids have different binding characteristics to hydroxyapatite. Such characteristics provide a basis commonly used for fractionating nucleic acids. Hydroxyapatite is commerically available [e.g., BioRad Laboratories, Hercules, CA (USA)]. The fraction of genomic DNA containing sequences with a particular degree of repetition, from the highest copy-number to single-copy, can be obtained by denaturing genomic DNA, allowing it to reassociate under appropriate conditions to a particular value of C_ot, followed by separation using hydroxyapatite. The single- and double-stranded nucleic acid can also be discriminated by use of S1 nuclease. Such techniques and the concept of C_ot are explained in Britten et al., "Analysis of Repeating DNA Sequences by Reassociation", in Methods in Enzymology, 29: 363-418 (1974).

Reaction with immobilized nucleic acid. Removal of particular sequences can also be accomplished by attaching single-stranded "absorbing" nucleic acid sequences to a solid support. Single-stranded source nucleic acid is hybridized to the immobilized nucleic acid. After the hybridization, the unbound sequences are collected and used in CGH. For example, human genomic DNA can be used to absorb repetitive sequences from the subject nucleic acids. One such method is described by Brison et al., "General Method for Cloning Amplified DNA by Differential Screening with Genomic Probes," Molecular and Cellular Biology, 2: 578-587 (1982). Briefly, minimally sheared human genomic DNA is bound to diazonium cellulose or a like support. The source DNA, appropriately cut into fragments, is hybridized against the immobilized DNA to C_ot

values in the range of about 1 to 100. The preferred stringency of the hybridization conditions may vary depending on the base composition of the DNA.

Prehybridization. Blocking of repeat sequence binding sites in the reference genome by hybridization with unlabeled complementary sequences will prevent binding of labeled sequences in the subject nucleic acids that have the potential to bind to those sites. For example, hybridization with unlabeled genomic DNA will render the high-copy repetitive sequences in the reference genome double-stranded. Labeled copies of such sequences in the subject nucleic acids will not be able to bind when they are subsequently applied.

In practice, several mechanisms can be combined to produce the desired contrast and sensitivity.

Labeling the Nucleic Acid Fragments of the Subject Nucleic Acids

There are many techniques available for labeling single- and double-stranded nucleic acid fragments of the subject nucleic acids. They include incorporation of radioactive labels, e.g. Harper et al. Chromosoma, 83: 431-439 (1984); direct attachment of fluorochromes or enzymes, e.g. Smith et al., Nuc. Acids Res., 13: 2399-2412 (1985), and Connolly et al., Nuc. Acids Res., 13: 4485-4502 (1985); and various chemical modifications of the nucleic acid fragments that render them detectable immunochemically or by other affinity reactions, e.g. Tchen et al., "Chemically Modified Nucleic Acids as Immunodetectable Probes in Hybridization Experiments," PNAS, 81: 3466-3470 (1984); Richardson et al., "Biotin and Fluorescent Labeling of RNA Using T4 RNA Ligase," Nuc. Acids Res., 11: 6167-6184 (1983); Langer et al., "Enzymatic Synthesis of Biotin-Labeled Polynucleotides: Novel Nucleic Acid Affinity Probes," PNAS, 78: 6633-6637 (1981); Brigati et al., "Detection of Viral Genomes in Cultured Cells and Paraffin-Embedded Tissue Sections Using Biotin-Labeled Hybridization Probes," Virol., 126: 32-50 (1983); Broker et al., "Electron Microscopic Visualization of tRNA Genes with Ferritin-Avidin: Biotin Labels," Nuc. Acids Res., 5: 363-384

(1978); Bayer et al., "The Use of the Avidin Biotin Complex as a Tool in Molecular Biology," Methods of Biochem. Analysis, 26: 1-45 (1980); Kuhlmann, Immunoenzyme Techniques in Cytochemistry (Weinheim, Basel, 1984). Langer-Safer et al., PNAS (USA), 79: 4381 (1982): Landegent et al., Exp. Cell Res., 153: 61 (1984); and Hopman et al., Exp. Cell Res., 169: 357 (1987). Thus, as indicated, a wide variety of direct and/or indirect means are available to enable visualization of the subject nucleic sequences that have hybridized to the reference genome. Suitable visualizing means include various ligands, radionuclides, fluorochromes and other fluorescers, chemiluminescers, enzyme substates or co-factors, particles, dyes and the like. Some preferred exemplary labeling means include those wherein the probe fragments are biotinylated, modified with N-acetoxy-N-2-acetylaminofluorene, modified with fluorescein isothiocyanate or other fluorochromes, modified with mercury/TNP ligand, sulfonated, digoxigeninated or contain T-T dimers.

A preferred method of labeling is tailing by terminal transferase labeling. Another preferred method is random priming with mixed sequence primers followed by polymerase extension. This has the additional feature of amplifying the amount of subject DNA, if several cycles are used, which is useful when only a small amount of DNA was originally obtained from the subject cell or cell population.

The key feature of labeling is that the subject nucleic acid fragments bound to the reference spread be detectable. In some cases, an intrinsic feature of the subject nucleic acid, rather than an added feature, can be exploited for this purpose. For example, antibodies that specifically recognize RNA/DNA duplexes have been demonstrated to have the ability to recognize probes made from RNA that are bound to DNA targets [Rudkin and Stollar, Nature, 265:472-473 (1977)]. The RNA used is unmodified. Nucleic acid fragments can be extended by adding "tails" of modified nucleotides or particular normal nucleotides. When a normal nucleotide tail is used, a second hybridization with nucleic acid complementary to the tail and containing fluorochromes,

enzymes, radioactivity, modified bases, among other labeling means, allows detection of the bound nucleic acid fragments. Such a system is commercially available from Enzo Biochem [Biobridge Labeling System; Enzo Biochem Inc., New York, N.Y. (USA)].

Another example of a means to visualize the bound nucleic acid fragments wherein the nucleic acid sequences do not directly carry some modified constituent is the use of antibodies to thymidine dimers. Nakane et al., ACTA Histochem. Cytochem., 20 (2):229 (1987), illustrate such a method wherein thymine-thymine dimerized DNA (T-T DNA) was used as a marker for *in situ* hybridization. The hybridized T-T DNA was detected immunohistochemically using rabbit anti-T-T DNA antibody.

All of the labeling techniques disclosed in the above references may be preferred under particular circumstances. Further, any labeling techniques known to those in the art would be useful to label the subject nucleic acids in of this invention. Several factors govern the choice of labeling means, including the effect of the label on the rate of hybridization and binding of the nucleic acid fragments to the chromosomal DNA, the accessibility of the bound nucleic acid fragments to labeling moieties applied after initial hybridization, the mutual compatibility of the labeling moieties, the nature and intensity of the signal generated by the label, the expense and ease in which the label is applied, and the like.

Several different subject nucleic acids, each labeled by a different method, can be used simultaneously. The binding of different nucleic acids can thereby be distinguished, for example, by different colors.

In Situ Hybridization.

Application of the subject nucleic acids to the reference chromosome spreads is accomplished by standard *in situ* hybridization techniques. Several excellent guides to the technique are available, e.g., Gall and Pardue, "Nucleic Acid Hybridization in Cytological Preparations," Methods in

Enzymology, 21: 470-480 (1981); Henderson, "Cytological Hybridization to Mammalian Chromosomes," International Review of Cytology, 76: 1-46 (1982); and Angerer et al., "in situ Hybridization to Cellular RNAs," in Genetic Engineering: Principles and Methods, Setlow and Hollaender, Eds., Vol. 7, pgs. 43-65 (Plenum Press, New York, 1985).

Generally in situ hybridization comprises the following major steps: (1) fixation of tissue or biological structure to be examined, (2) prehybridization treatment of the biological structure to increase accessibility of target DNA, and to reduce nonspecific binding, (3) hybridization of the mixture of nucleic acids to the nucleic acid in the biological structure or tissue; (4) posthybridization washes to remove nucleic acid fragments not bound in the hybridization and (5) detection of the hybridized nucleic acid fragments. The reagents used in each of these steps and their conditions of use vary depending on the particular situation.

Under the conditions of hybridization wherein human genomic DNA is used as an agent to block the hybridization capacity of the repetitive sequences, the preferred size range of the nucleic acid fragments is from about 200 bases to about 1000 bases, more preferably about 400 to 800 bases for double-stranded, nick-translated nucleic acids and about 200 to 600 bases for single-stranded or PCR adapter-linker amplified nucleic acids.

Example 1 provides details of a preferred hybridization protocol. Basically the same hybridization protocols as used for chromosome-specific painting as described in Pinkel et al., PNAS (USA), 85: 9138-9142 (1988) and in EP Pub. No. 430,402 (published June 5, 1991) are adapted for use in CGH.

The following representative examples of performing CGH methods of this invention are for purposes of illustration only and are not meant to limit the invention in any way.

Example 1

DNA from Breast Cancer Lines Hybridized to Normal Metaphase Spreads

In this Example, methods of this invention to analyse genomes by Comparative Genomic Hybridization (CGH) are exemplified by hybridizations of breast cancer cell lines to normal metaphase spreads. The target metaphase spreads were pre-hybridized with unlabeled human placental DNA to block the high copy repeat sequences. In this representative example, the hybridization mixture containing the extracted labeled DNA from the cell lines contained unlabeled, repeat-enriched Cot-1 blocking DNA [obtained from Bethesda Research Laboratories (BRL), Gaithersburg, MD (USA].

The experiments outlined below include in the hybridization mixture for the subject genomes, that is, the breast cancer cell line DNAs, chromosome-specific repeat sequence probes and chromosome-specific painting probes. Those probes labeled with biotin were included as an adjunct for identifying chromosomes in the metaphase preparations. The experiments were first performed without those chromosome-specific probes. Then each chromosome of interest was measured to determine its length which was considered along with other factors to determine its probable identity. The chromosome-specific probes were then used in the hybridization mixture to confirm the identity of the chromosome of interest. However, such probes are not necessary as the chromosomes could have been identified by the DAPI banding of the counterstain or by other chemical staining, such as staining with quinacrine, by a skilled cytogeneticist.

Cell Lines and Isolation of DNA:

Six established breast cancer cell lines: BT-474, SK-BR-3, MCF-7, MDA-MB-361, MDA-MB-468 and T-47D were obtained from the American Type Culture Collection [Rockville, Maryland (USA)]. The breast cancer cell line 600MPE cell line was kindly provided by Dr. Helene S. Smith [Geraldine Brush Cancer Research Center, San Francisco, CA (USA)]. Cell lines were grown until they became confluent. Cells were then

trypsinized, pelleted by centrifugation at 1500 RPM for 5 minutes and washed twice in phosphate buffered saline. The DNA was then isolated as described by Sambrook et al., Molecular Cloning: A Laboratory Manual, Vol. 2: 9.16-9.19 [Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (USA) 1989].

Details concerning the established human breast cancer cell lines used herein are as follows:

BT-474	-	Originated from a human primary cancer; obtained from the ATCC, catalog # HTB 20;
SK-BR-3	-	Originated from a human metastatic breast adenocarcinoma derived from a pleural effusion; obtained from the ATCC catalog # HTB 30;
MDA-MB-361	-	Originated as a metastatic tumor to the brain; obtained from the ATCC, catalog # HTB 27;
MCF-7	-	Originated from a human metastatic pleural effusion; obtained from the ATCC, catalog # HTB 22;
T-47D	-	Originated as a human metastatic pleural effusion; obtained from the ATCC catalog # HTB 133;
600MPE	-	Originated as a human metastatic pleural effusion; kindly provided by Dr. Helene S. Smith [Geraldine Brush Cancer Research Center, San Francisco, CA (USA)]; and
MDA-MB-468	-	Originated as a metastatic pleural effusion; obtained from the ATCC, catalog # HTB 132.

Preparation of Normal Lymphocyte Metaphases:

Normal peripheral blood lymphocytes were stimulated by PHA, synchronized by methotrexate treatment and blocked in metaphase using 0.05 ug/ml colcemid. Cells were then centrifuged, washed and incubated in 75 mM KCl at 37°C for 15 minutes. Cells were then fixed in methanol:acetic acid (3:1) and dropped onto slides. The slides were stored under nitrogen at -20°C.

DNA Labeling:

Cell line DNAs were labeled with digoxigenin-11-dUTP using nick translation [Rigby et al., J. Mol. Biol., 113: 237 (1977); Sambrook et al., supra]. The optimal size of the probe fragments after nick translation and before denaturing was 400-800 bps. As indicated above, chromosome-specific probes were used in dual-color hybridizations to verify the identification of chromosomes of interest in the metaphase spreads. Representative examples of such chromosome-specific reference probes labeled with biotin-14-dATP include the following:

1) a chromosome-specific painting probe for chromosome 20 prepared by the PCR adapter-linker method as described in PCT/US90/00434 published August 9, 1990;

2) a chromosome 17 peri-centromeric repeat probe (cosmid cK17.10) isolated by Anne Kallioniemi from a chromosome 17 cosmid library from Los Alamos National Laboratory [Albuquerque, New Mexico (USA)]; an equivalent chromosome-specific repeat sequence probe for chromosome 17 is commercially available from Oncor [Gaithersburg, MD (USA)]; and

3) an alpha satellite repeat probe specific for the centromeric region of chromosome 8 [kindly provided by Dr. Heinz-Ulrich G. Weier; University of California Medical Center, Lab for Cell Analysis, San Francisco, CA (USA)]; that probe was generated by Dr. Weier using PCR with primers WA1 and WA2 as described in Weier et al., Hum. Genet., 87: 489-494 (1991).

Ones skilled in the art recognize that there are many other equivalent probes available that could be used for the confirmation purposes described. For example, whole chromosome painting probes are now available for all the human chromosomes [Collins et al., Genomics, 11: 997-1006 (1991)]. Also available are repeat sequence probes that hybridize intensely and specifically to selected chromosomes [Trask et al., Hum. Genet., 78: 251 (1988) and references cited therein].

Pretreatment and Prehybridization of Slides:

Lymphocyte metaphase preparations were first denatured in 70% formamide/2XSSC (1XSSC is 0.15 M NaCl, 0.015 M NaCitrate), pH 7, at 70°C for 2 minutes and dehydrated in a sequence of 70%, 85% and 100% ethanol. The slides were then air dried and treated with 10 ug/50 ml Proteinase K [Boehringer Mannheim GmbH, Indianapolis IN (USA)] for 7.5 minutes at 37°C in a buffer containing 20 mM Tris and 2 mM $CaCl_2$ (pH 7.5). Ethanol dehydration was then done as described above, and the slides were prehybridized with ten ul of a hybridization mixture, consisting of 20 ug unlabeled human placental DNA [obtained from Sigma, St. Louis, MO (USA); size of the fragments is 200-700 bps] in 50% formamide, 10% dextran sulphate and 2XSSC (pH 7) for 60 minutes at 37°C. Before the prehybridization mixture was applied to the slides, it was denatured in a 70°C water bath for 5 minutes. After prehybridization, the slides were washed once in 2XSSC and dehydrated with ethanol as described above.

Hybridization:

Five ug of unlabeled, repeat-enriched Cot-1 blocking DNA [BRL, Gaithersburg, MD (USA)] and 60 ng of digoxigenin labeled cell line DNA and 20-60 ng of biotin-labeled reference probes (for verification of chromosome identification) were mixed together and 1/10 vol of 3M Na-acetate was added. DNA was precipitated by adding 2 volumes of 100% ethanol followed by centrifugation in a microcentrifuge for 30 minutes at 15,000 RPM. Ethanol was removed and the tubes were allowed to dry until all visible ethanol had evaporated. Ten ul of hybridization buffer consisting of 50% formamide, 10% dextran sulphate and 2XSSC (pH 7) was then added, followed by careful mixing. DNAs in the hybridization buffer were then denatured for 5 minutes at 70°C followed by a 60 minute renaturation at 37°C. The hybridization mixture was then added to the prehybridized lymphocyte metaphase slides. Hybridization was carried out under a coverslip in a moist chamber for 3-4 days at 37°C.

Immunofluorescent Probe Detection:

The slides were washed three times in 50% formamide/ 2XSSC, pH 7, twice in 2XSSC and once in 0.1XSSC for 10 minutes each at 45°C. After washing, the slides were immunocytochemically stained at room temperature in three steps (30-45 minutes each). Before the first immunocytochemical staining, the slides were preblocked in 1% BSA/4XSSC for 5 minutes. The first staining step consisted of 2 ug/ml Texas Red-Avidin [Vector Laboratories, Inc., Burlingame, CA (USA)] in 1% BSA/4XSSC. The slides were then washed in 4XSSC, 4XSSC/0.1% Triton X-100, 4XSSC, and PN (a mixture of 0.1 M NaH_2PO_4 and 0.1 M Na_2HPO_4, pH 8, and 0.1% Nonidet P-40) for 10 minutes each and preblocked with PNM (5% Carnation dry milk, 0.02% Na-azide in PN buffer) for 5 minutes. The second antibody incubation consisted of 2 ug/ml FITC-conjugated sheep anti-digoxigenin [Boehringer Mannheim GmBH, Indianapolis, IN (USA)] and 5 ug/ml anti-avidin [Vector Laboratories, Burlingame, CA (USA)] in PNM followed by three PN washes, 10 minutes each. After the PNM block, the third immunochemical staining was done using rabbit anti-sheep FITC antibody (1:50 dilution) (Vector Laboratories) and 2 ug/ml Texas Red-Avidin in PNM. After three PN washes, nuclei were counterstained with 0.8 uM 4,5-diamidino-2-phenylindole (DAPI) in an antifade solution.

Fluorescence Microscopy and Interpretation of Results:

A Nikon fluorescence microscope [Nikon Inc., Garden City, NY (USA)] equipped with a double band pass filter [Chroma Technology, Brattleboro, VT (USA)] and a 100X objective was used for simultaneous visualization of the FITC and Texas Red signals. Hybridization of the breast cancer cell line DNAs was seen as a more or less uniform faint green background staining of all metaphase chromosomes with the exception of the Y-chromosome. As the breast cancer cell lines are of course of female origin, they did not contain Y chromosomal DNA. The absence of said green staining of the Y chromosome of the metaphase spread is exemplary of the manner in which a cytogenetically significant deletion would be

visualized. The absence of the Y chromosome in the breast cancer cell line DNA was detected, as would a cytogenetically significant deletion, by the hybridization wherein the Y chromosome of the reference spread was only stained by the DAPI counterstain. Using a fluorescence microscope, amplified sequences can be seen as bright green dots or bands along the chromosome arms.

To facilitate the display of the results and to improve the sensitivity of detecting small differences in fluorescence intensity, a digital image analysis system (QUIPS) was used. QUIPS (an acronym for quantitative image processing system) is an automated image analysis system based on a standard Nikon Microphot SA [Nikon Inc., Garden City, NY (USA)] fluorescence microscope equipped with an automated stage, focus control and filterwheel [Ludl Electronic Products Ltd., Hawthorne, NY (USA)]. The filterwheel is mounted in the fluorescence excitation path of the microscope for selection of the excitation wavelength. Special filters [Chroma Technology, Brattleboro, VT (USA)] in the dichroic block allow excitation of multiple dyes without image registration shift. The microscope has two camera ports, one of which has an intensified CCD camera [Quantex Corp., Sunnyvale, CA (USA)] for sensitive high-speed video image display which is used for finding interesting areas on a slide as well as for focusing. The other camera port has a cooled CCD camera [model 200 by Photometrics Ltd., Tucson, AZ (USA)] which is used for the actual image acquisition at high resolution and sensitivity.

The cooled CCD camera is interfaced to a SUN 4/330 workstation [SUN Microsystems Inc., Mountain View, CA (USA)] through a VME bus. The entire acquisition of multicolor images is controlled using an image processing software package SCIL-Image [Delft Centre for Image Processing, Delft, Netherlands]. Other options for controlling the cameras, stage, focus and filterwheel as well as special programs for the acquisition and display of multicolor images were developed at the Division of Molecular Cytometry [University of California, Medical Center; San Francisco, CA (USA)] based on the SCIL-Image package.

To display the results of the comparative hybridization, two or three consecutive images were acquired (DAPI, FITC and Texas Red) and superimposed. The FITC image was displayed after using the thresholding and contrast enhancement options of the SCIL-Image software. Exercising such options reduces the overall chromosomal fluorescence to make amplified sequences more readily visible. For example, using thresholding and contrast stretching, it was possible to enhance the contrast and quantification between the faint green background staining and staining originating from the amplified sequences in the cell lines. Alternatively, to facilitate the detection of deletions, it is possible to increase the overall chromosomal fluorescence and make areas of reduced fluorescence appear darker. The red color was used for reference probes to help in the identification of chromosomes.

After identification of the chromosomes based on the use of reference probes in a dual-color hybridization, a site of amplification was localized by fractional length measurements along the chromosome arm (fractional length = distance of the hybridization signal from the p-telomere divided by the total length of the chromosome). The band location of the signal was then approximated from the fractional length estimate based on the ISCN 1985 idiograms [Harnden and Klinger, An International System for Cytogenetic Nomenclature, Karger Ag, Basel, Switzerland (1985)].

Results:

The results from the hybridizations are compiled in Table 2 along with other information known about the cell lines. Amplification at 17q12 (erbB-2 locus) and approximately 8q24 (MYC locus) was seen in lines showing amplification of erbB-2 and MYC whenever the level of amplification was greater than about five- to ten-fold using this CGH method. In addition, amplification of several megabase wide regions was seen in three cell lines at 17q22-23 and in three lines at 20qter; those amplifications were previously unknown sites of amplification and were not

expected from other studies. For example, as indicated in Table 2, the BT-474 cell line is known to have a 13-fold c-erbB-2 amplification; CGH revealed amplified sequences at the following loci: 17q12 (the erbB-2 locus); 17q22-q23 and 20q13-ter. The latter two sites were previously unrecognized sites of amplification in that cell line.

All lines showing amplification showed amplification at more than one site. Evidence for co-amplification may be clinically important since co-amplification has been observed previously [van de Vijver et al., Mol. Cell Biol. 7: 2019-2023 (1987); Saint-Ruf et al., Oncogene, 6: 403-406 (1991)], and is sometimes associated with poor prognosis [Borg et al., Br. J. Cancer, 63: 136-142 (1991)]. Amplification at 17q22-23 has also been seen using probe DNA from primary tumors.

TABLE 2

Results of Testing Breast Cancer Cell Lines for Amplification

Cell Line	Origin	Growth rate; -Td	Hormone receptor E/P	Known amplification (level)	Amplification detected by CGH
BT-474	Primary Cancer	48-96 hr	+/-	erbB-2 (13X)	17q12 (erbB-2), 17q22-23, 20qter
SK-BR-3	Pl. Effusion	?	?	erbB-2 (9X) MYC(10X)	17q12 (erbB-2), 8q21, 8q23-24.1 (MYC), 20qter
MDA-MB-361	Brain met.	<96 hr	-/+	erbB-2 (4X)	17q22-23
MCF-7	Pl. Effusion	<48 hr	+/+	erbB-2 (none)	17q22-23, 20qter
T-47D	Pl. Effusion	?	+/+	erbB-2 (none)	None
600MPE	Pl. Effusion	?	?	erbB-2 (none)	None
MDA-MB-468	Pl. Effusion	?	?	erbB-2 (none)	None

Example 2

Hybridizations with two different labeled subject DNAs as schematically outlined in Figures 5 and 6 were performed. One of the labeled subject DNAs hybridized was a cell line DNA as described in Example 1 and similarly labeled. The other labeled subject DNA was human genomic DNA labeled with biotin-14-dATP.

The protocols were essentially the same as in Example 1 except that no chromosome-specific reference probes were used, and the same amount of the labeled human DNA as the labeled cell line DNA, that is, 60 ng, was hybridized. Of course, reference probes could be added to the hybridization mixture, but they need to be differently labeled to be distinguishable.

The results showed the normal DNA with a red signal and the cell line DNA with a green signal. The green to red ratios were determined along each chromosome. Amplification was indicated by an area where the signal was predominantly green whereas deletions were indicated by more red signals than in other areas of the chromosomes.

Exemplary, CGH results using breast cancer cell line 600MPE DNA and normal human DNA were as follows. As indicated above, the hybridization was performed using 5 ug Cot-1 DNA, 60 ng of digoxigenin labeled 600MPE cell line DNA, and 60 ng of biotinylated normal human genomic DNA. The 600MPE DNA was detected with FITC (green) and the genomic DNA with Texas Red-Avidin (red).

The 600MPE breast cancer cell line, the karyotype for which was published by Smith et al., JNCI, 78: 611-615 (1987), contains one normal chromosome 1 and three marker chromosomes with chromosome 1 material in them: t(1q:13q), 1p(p22) and inv(1)(p36q21). Thus, the cell line is disomic for the p-telomere-p22, trisomic for p22-centromere and tetrasomic for the q-arm of chromosome 1.

The comparative genomic hybridizations of this example apparently identified three different regions on chromosome 1 that could be separated according to the intensities of green and red colors. The q-arm of chromosome

1 had the highest intensity of green color (tumor DNA). The region from band p22 to the centromere was the second brightest in green, and the area from the p-telomere to band p22 had the highest intensity of red color (normal DNA). Those hybridization results were consistent with the traditional cytogenetic analyses of that cell line stated immediately above.

However, further studies with CGH, as presented in Example 3, indicated that the CGH analysis of this example, as well as the published karyotype, were partially in error. The CGH analysis of Example 3 motivated additional confirmatory experiments, as described therein, leading to correction of the original CGH results and the published karyotype.

Example 3

Copy Number Karyotypes of Tumor DNA

In the representative experiments of CGH in this example, biotinylated total tumor DNA (cell line and primary tumor DNA) and digoxigenin-labeled normal human genomic DNA are simultaneously hybridized to normal human metaphase spreads in the presence of unlabeled blocking DNA containing high-copy repetitive sequences, specifically unlabeled Cot-1 blocking DNA [BRL, Gaithersburg, MD (USA)]. The following paragraphs detail the procedures used for the representative CGH experiments of this example.

DNA Labeling:

DNAs used in this example were labeled essentially as shown above in Example 1. DNAs were labeled with biotin-14-dATP or digoxigenin-11-dUTP by nick translation [Rigby et al., supra; Sambrook et al., supra]. The optimal size for double stranded probe fragments after labeling was 600-1000 bp.

Pretreatment of Metaphase Spreads:

Lymphocyte metaphase preparations were denatured, dehydrated and air dried, treated with Proteinase K and dehydrated again as described in Example 1.

Comparative Genomic Hybridization:

Sixty ng of biotinylated test DNA, 60 ng of digoxigenin-labeled normal DNA and 5 μg of unlabeled Cot-1 DNA (BRL) were ethanol precipitated and dissolved in 10 μl of 50% formamide, 10% dextran sulfate, 2xSSC, pH 7. The probe mixture was denatured at 70°C for 5 minutes, allowed to reanneal at 37°C for 60 minutes and hybridized to normal male metaphase chromosomes for 3-4 days at 37°C.

Immunofluorescent Probe Detection:

The slides were washed as described above in Example 1, and immunocytochemically stained at room temperature in three thirty-minute steps: (I) 5 μg/ml FITC-Avidin [Vector Laboratories, Inc., Burlingame, CA (USA)] and 2 μg/ml anti-digoxigenin-Rhodamine (Boehringer Mannheim GMbH); (II) 5 μg/ml anti-avidin (Vector Laboratories); and (III) 5 μg/ml FITC-avidin. Nuclei were counterstained with 0.8 μM 4,5-diamino-2-phenylindole (DAPI) in antifade solution. A Zeiss fluorescence microscope equipped with a double band pass filter [Chroma Technology, Brattleboro, VT (USA)] was used for simultaneous visualization of FITC and rhodamine signals.

Digital Image Analysis System and Fluorescence Ratio Profiles

The QUIPS system essentially a described above in Example 1 was used to analyse quantitatively the fluorescence signals. Fluorescence ratio profiles along the chromosomes were extracted using WOOLZ software package [developed at MRC, Edinburgh, Scotland] as follows: the DAPI image is used to set the morphological boundary of each chromosome by thresholding. The chromosome outline is smoothed by a n number of opening and closing operations, a modified Hilditch skeleton is calculated and taken to represent the medial axis of the chromosome. The DAPI image is expanded outwards in all directions until the intensity field levels off (when background is reached) or begins to rise (due to an adjacent chromosome). The intensity profile of each image along the medial axis and within the expanded DAPI image is then calculated by summing the green and red fluorescence pixel

values along the sequence of lines perpendicular to and spaced at unit distance along the medial axis. Modal green and red intensity values corresponding to the expanded DAPI image are taken to represent the background fluorescence and used as the intensity origin.

<u>Cell Lines</u>:

5637	-	Originated from a human primary bladder carcinoma; obtained from ATCC, catalog # HTB 9
SK-BR-3	-	Originated from a human metastatic breast adenocarcinoma, derived from a pleural effusion; obtained from the ATCC, catalog # HTB 30
Colo 205	-	Originated from a human colon adenocarcinoma; obtained from the ATCC, catalog # CCL 222
NCI-H508	-	Originated from a human cecum adenocarcinoma; obtained from the ATCC, catalog # CCL 253
SW480	-	Originated from a human colon adenocarcinoma; obtained from the ATCC, catalog # CCL 228
SW620	-	Originated from a human lymph node metatasis of a colon adenocarcinoma; obtained from the ATCC, catalog # CCL 227
WiDr	-	Originated from a human colon adenocarcinoma; obtained from the ATCC, catalog # CCL 218
SK-N-MC	-	Originated from a human neuroblastoma (metastasis to supra-orbital area); obtained from the ATCC, catalog # HTB 10
CaLu3	-	Originated from a human lung adenocarcinoma, derived from a pleural effusion; obtained from the ATCC, catalog # HTB 55
CaLu6	-	Originated from a human anaplastic carcinoma, probably lung; obtained from the ATCC, catalog # HTB 56
NCI-H69	-	Originated from a human small cell lung carcinoma; obtained from the ATCC, catalog # HTB 119

COLO 320HSR	-	Originated from a human colon adenocarcinoma; obtained from the ATCC, catalog # 220.1
600 PE	-	Originated from a human breast carcinoma; obtained from Dr. Helene Smith and Dr. Ling Chen [Geraldine Brush Cancer Research Center, San Francisco, CA (USA)]. This is the same as the 600 MPE cell line described in Examples 1 and 2.
BT-20	-	Originated from a human breast carcinoma; obtained from ATCC, catalog # HTB 19

The following are five fibroblast cell lines with total chromosomal number and X chromosomal number in parentheses, which were obtained from the NIGMS repository [Camden, NJ (USA)]:

GM01723 (45,XO)
GM08399 (46,XX)
GM04626 (47,XXX)
GM01415E (48,XXXX)
GM05009B (49,XXXXX).

Results and Discussion:

Demonstrated herein is CGH's capability of detecting and mapping relative DNA sequence copy number between genomes. A comparison of DNAs from malignant and normal cells permits the generation of a "copy number karyotype" for a tumor, thereby identifying regions of gain or loss of DNA.

Demonstrated is the use of dual color fluorescence in situ hybridization of differently labeled DNAs from a subject tumor genome and a normal human genome to a normal human metaphase spread to map DNA sequence copy number throughout the tumor genome being tested. Regions of gain or loss of DNA sequences, such as deletions, duplications or amplifications, are seen as changes in the ratio of the intensities of the two fluorochromes (used in this representative example) along the target chromosomes. Analysis of tumor cell lines and primary bladder tumors identified 16 different regions of amplification, many in loci not previously known to be amplified. Those results are shown in Table 3 below.

The tumor DNA is detected with the green fluorescing FITC-avidin, and the normal DNA with the red fluorescing rhodamine anti-digoxigenin. The relative amounts of tumor and normal DNA bound at a given chromosomal locus are dependent on the relative abundance of those sequences in the two DNA samples, and can be quantitated by measurement of the ratio of green to red fluorescence. The normal DNA in this example serves as a control for local variations in the ability to hybridize to target chromosomes. Thus, gene amplification or chromosomal duplication in the tumor DNA produces an elevated green-to-red ratio, and deletions or chromosomal loss cause a reduced ratio. The Cot-1 DNA included in the hybridization inhibits binding of the labeled DNAs to the centromeric and heterochromatic regions so those regions are excluded from the analysis.

The fluorescence signals were quantitatively analyzed by means of a digital image analysis system as described above. A software program integrated the green and red fluorescence intensities in strips orthogonal to the chromosomal axis, subtracted local background, and calculated intensity profiles for both colors and the green-to-red ratio along the chromosomes.

The ability of CGH to quantitate changes in sequence copy number that affect an entire chromosome was tested with the above-listed five fibroblast cell lines having 1 to 5 copies of the X chromosome and two copies of each autosome. Hybridization of DNA from the 45,XO cell line (in green) together with normal female DNA (in red) resulted in a uniform green-red staining of the autosomes whereas the X chromosome appeared more red. (The reference spread as indicated above was of normal male chromosomes. Some faint staining of a small part of the Y chromosome was the result of the binding of homologous sequences in the pseudo-autosomal region.)

Hybridizations with DNA from cell lines carrying 2, 3, 4 or 5 copies of the X chromosome resulted in an increasingly strong green fluorescence from the X chromosome in relation to the autosomes. The average green-to-red fluorescence ratio of the X chromosome (Figure 7), when

normalized to the average ratio for the autosomes within the same metaphase spread, increased linearly with the increasing number of X chromosomes [correlation coefficient (r) = 0.978]. Thus, CGH can quantitatively distinguish a change of plus or minus one copy of a chromosome at least up to 4 copies.

Experiments showed that CGH could generate a complete copy number karyotype for a near-diploid breast cancer cell line, 600PE. According to the published karyotype for 600PE [Smith et al., JNCI, 78: 611 (1987)], 600PE is near-diploid with five marker chromosomes having four copies of the q-arm of chromosome 1, monosomy 16, and deletions of 9p, 11q and 17p. CGH using biotinylated 600PE DNA (in green) and normal digoxigenin-labeled DNA (in red) revealed the following relative copy number changes: gain of 1q and loss of 9p, 16q, 17p and distal 11q. The green-to-red ratio profiles for those aberrant chromosomes are shown in Figure 8. Only the q-arm of chromosome 16 showed decreased relative copy number suggesting that 16p was not deleted. That observation was subsequently confirmed by fluorescence in situ hybridization (FISH) to 600PE interphase cells using cosmid probes for the p- and q-arms of chromosome 16 [16p and 16q cosmid probes provided by Los Alamos National Laboratory, Los Alamos, NM (USA)]; two signals per nucleus for the 16p cosmid probe and one for the 16q cosmid probe permitted calibration of a green-to-red ratio of 1.0 as indicating two copies of a sequence.

Thus, if the absolute copy number of any point in the tumor genome is known, relative copy numbers can be converted to actual copy numbers at all loci. The CGH results differed from the originally published karyotype in the region of 16p and proximal 1p. That discrepancy was resolved by locus-specific chromosome-specific painting (FISH) that indicated that the components of one of the marker chromosomes had been misinterpreted by conventional cytogenetic analysis.

CGH with DNAs from two fibroblast cell lines [GM05877 and GM01142A from the NIGMS repository] detected small interstitial deletions around the RB-1 locus in 13q--del(13) (pter > q14.1::q21.2 > qter) and del(13) (pter >

q14.1::q22.1 > qter). On the basis of the CGH analysis and measurement of the deletion size as a fraction of the length of chromosome 13 [total length 111 megabases (Mb)], those deletions were estimated to span about 10 and 20 Mb, respectively. Thus, it is possible that CGH can be used to screen DNA samples from solid tumors in order to identify large physical deletions that may uncover recessive mutant tumor suppressor genes.

CGH was evaluated for its ability to detect increased gene copy number with cell lines that contained previously reported amplification of oncogenes. Figure 9A shows CGH with DNA from a colon cancer cell line COLO 320HSR, known to contain more than a 50-fold amplification of a 300 kb region around the <u>myc</u> oncogene [Kinzku et al., <u>PNAS</u> (USA), <u>83</u>: 1031 (1986)]. The expected high green-to-red ratio at 8q24 corresponding to the location of <u>myc</u> is clear. The height of the peak does not quantitatively reflect the level of amplification because the fluorescent signal spread over a region of the chromosome that is larger than the length of the amplicon. That is apparently a result of the complex organization of the target DNA in the denatured chromosomes.

The eight-fold amplification of the <u>erb</u>B2 oncogene in the SK-BR-3 breast cancer cell line also was detectable with CGH as a hybridization signal at 17q12 (Table 3). High level amplifications such as those also could be detected in single color-hybridizations with the use of only labeled tumor DNA.

Cytogenetic and molecular studies of primary tumors and cell lines often reveal homogeneously staining regions and double minute chromosomes that do not involve known oncogenes [Saint-Ruf et al., <u>Genes Chrom. Cancer., 2</u>: 18 (1990); Bruderlein et al., <u>Genes Chrom. Cancer, 2</u>: 63 (1990)]. CGH allows straightforward detection and mapping of such sequences. Table 3 contains a summary of the analysis with CGH of 11 cancer cell lines. Data in Table 3 is based on the visual inspection of a large number of metaphase spreads and on detailed digital image analysis of four to six metaphases for each sample.

TABLE 3

Mapping of amplified sequences in established cancer cell lines and primary tumors by CGH

Specimen	Origin	Amplif. by CGH*	Cytogenetic evidence of gene amplif.+
Cell lines:			
5637	Bladder	3p25, 6p22	DM
SK-BR-3	Breast	8q24 (myc), 8q21, 17q12 (erbB2), 20q13	
Colo 205	Colorectal	6p21, 6q24	
NCI-H508	Colorectal	14q12-13	DM
SW480	Colorectal	8q24 (myc)	DM
SW620	Colorectal	16q21-23	HSR
WiDr	Colorectal	8q23-24 (myc)	
SK-N-MC	Neuroblastoma	8q24 (myc)	DM
CaLu3	Small cell lung	8p12-21, 8qtel, 17q12 (erbB2)	HSR
CaLu6	Small cell lung	13q32-34	
NCI-H69	Small cell lung	2p24 (N-myc), 2p21, 2q21	
Primary tumors:			
UR140	Bladder carcinoma	16q21-22	
UR145	Bladder carcinoma	6p22	

* The oncogene most likely involved in this amplification is shown in parentheses.
\+ Cytogenetic information based on the ATCC Catalogue of Cell Lines & Hybridomas (1992).
DM = double minute chromosomes, HSR = homogeneously staining regions.

Sixteen amplified loci were mapped, many at regions of the genome where amplification had not previously been suspected. Thus, a large variety of genes may be amplified during cancer initiation and progression. In five of the 11

cell lines, more than one locus was amplified. Two or three separate loci on the same chromosome were amplified in four cell lines, which suggests a spatial clustering of chromosomal locations that undergo DNA amplification (Table 3 and Figure 9B).

CGH was also applied to identify and map amplified DNA sequences in uncultured primary bladder tumors. Of the seven tumors tested, two showed evidence of DNA amplification but the loci were not the same (Table 3). Thus, a number of previously unsuspected genomic regions that might contain genes important for cancer progression have been identified by CGH. Further studies will elucidate which of those loci contain novel oncogenes and which represent coincidental, random DNA amplification characteristic of genomic instability.

The detection and mapping of unknown amplified sequences that typically span several hundred kilobases (kb) to a few Mb demonstrated the usefulness of CGH for rapid identification of regions of the genome that may contain oncogenes. Analogously, detection of deletions may facilitate identification of regions that contain tumor suppressor genes.

Further studies are necessary to establish to what extent allelic losses in tumors are caused by physical deletions. In clinical specimens, the detection of small copy number differences is more difficult than with cell lines because of the admixture of DNA from contaminating normal cells and because of intratumor heterogeneity. As indicated above, using PCR to prepare tumor DNA from a small number of tumor cells (as a tumor clonal sub-population) may assist in resolving that problem. Like RFLP, CGH emphasizes the detection of aberrations that are homogeneous in a cell population and averages those that are heterogeneous.

At the current stage of development of CGH, sensitivity is primarily limited by the granularity of the hybridization signals in the metaphase chromosomes. Further improvements in sensitivity will be achieved by optimization of the probe concentration and labeling, and by the averaging

of the green-to-red fluorescence ratios from several metaphase spreads.

The descriptions of the foregoing embodiments of the invention have been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and obviously many modifications and variations are possible in light of the above teachings. The embodiments were chosen and described in order to best explain the principles of the invention and its practical application to enable thereby others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto. All references cited herein are incorporated by reference.

CLAIMS

1. A method of comparing copy numbers of different DNA sequences in a subject cell or cell population comprising the steps of:

a) extracting the DNA from the subject cell or from a number of cells of the subject cell population;

b) amplifying said extracted subject DNA, if necessary;

c) labeling the subject DNA;

d) hybridizing said labeled subject DNA <u>in situ</u> to reference metaphase chromosomes after substantially removing from the labeled subject DNA those repetitive sequences that could bind to multiple loci in the reference metaphase chromosomes, and/or after blocking the binding sites for those repetitive sequences in the reference metaphase chromosomes by prehybridization with appropriate blocking nucleic acids, and/or blocking those repetitive sequences in the labeled subject DNA by prehybridization with appropriate blocking nucleic acid sequences, and/or including such blocking nucleic acid sequences for said repetitive sequences during said hybridization, wherein the DNA sequences in the labeled subject DNA that bind to single copy sequences in the reference metaphase chromosomes are substantially retained, and those single copy DNA sequences as well as their binding sites in the reference metaphase chromosomes remain substantially unblocked both before and during the hybridization;

e) rendering the bound, labeled subject DNA sequences visualizable, if necessary;

f) observing and/or measuring the intensity of the signal from the bound labeled subject DNA sequences as a function of position on the reference metaphase chromosomes; and

g) comparing the copy numbers of different DNA sequences of the subject DNA by comparing the signal intensities at different positions on the reference metaphase chromosomes, wherein the greater the signal intensity at a

given position, the greater the copy number of the sequences in the subject DNA that bind at that position.

2. A method according to Claim 1 wherein the copy number of a subject DNA sequence binding at one position in the reference metaphase chromosomes relative to the copy number of a sequence binding at another position is quantified by measuring the ratio of the signal intensities at the two locations.

3. A method according to Claim 1 further comprising the addition of an unlabeled nucleic acid to said hybridization mixture wherein said unlabeled nucleic acid has a sufficient number of nucleic acid sequences substantially complementary to the sequences in the reference metaphase chromosomes to prevent saturation of the binding sites in the reference metaphase chromosomes by the labeled subject DNA.

4. A method according to Claim 1 wherein the reference metaphase chromosomes are human and prehybridized with human genomic DNA or human genomic DNA enriched in high copy repetitive sequences, and wherein human genomic DNA and/or human genomic DNA enriched in high copy repetitive sequences are included in the hybridization.

5. A method according to Claim 1 wherein in step (f) and/or (g), an image analysis system is used.

6. A method according to Claim 1 wherein signal intensities significantly greater than average indicate the presence and locations on the reference metaphase chromosomes of sequences that are duplicated or amplified in the subject cell or cell population whereas signal intensities that are significantly less than average indicate the presence and locations of sequences that are deleted in the subject cell or cell population.

7. A method according to Claim 1 wherein the reference metaphase chromosomes are from an antenna cell line.

8. A method of comparing copy numbers of different RNA sequences in a subject cell or cell population comprising the steps of:

a) extracting the RNA from the subject cell or from a number of cells of the subject cell population;

b) amplifying said extracted subject RNA, if necessary;

c) labeling the subject RNA;

d) hybridizing said labeled subject RNA _in situ_ to reference metaphase chromosomes after substantially removing from the labeled subject RNA those repetitive sequences that could bind to multiple loci in the reference metaphase chromosomes, and/or after blocking the binding sites for those repetitive sequences in the reference metaphase chromosomes by prehybridization with appropriate blocking nucleic acids, and/or blocking those repetitive sequences in the labeled subject RNA by prehybridization with appropriate blocking nucleic acid sequences, and/or including such blocking nucleic acid sequences for said repetitive sequences during said hybridization, wherein the RNA sequences in the labeled subject RNA that bind to single copy sequences in the reference metaphase chromosomes are substantially retained, and those RNA sequences as well as their binding sites in the reference metaphase chromosomes remain substantially unblocked both before and during the hybridization;

e) rendering the bound, labeled subject RNA sequences visualizable, if necessary;

f) observing and/or measuring the intensities of the signals from the bound labeled subject RNA sequences as a function of position on the reference metaphase chromosomes; and

g) comparing the copy numbers of different RNA sequences of the subject RNA by comparing the signal intensities at different positions on the reference metaphase chromosomes, wherein the greater the signal intensity at a

given position, the greater the copy number of the sequences in the subject RNA that bind at that position.

9. A method of comparing copy numbers of different DNA sequences in one subject cell or cell population relative to copy numbers of substantially identical sequences in another cell or cell population, said method comprising the steps of:

a) extracting the DNA from both of the subject cells or cell populations;

b) amplifying said extracted subject DNAs, if necessary;

c) labeling each of the subject DNAs differently;

d) hybridizing said differently labeled subject DNAs <u>in situ</u> to reference metaphase chromosomes after substantially removing from the labeled DNAs those repetitive sequences that could bind to multiple loci in the reference metaphase chromosomes, and/or after blocking the binding sites for those repetitive sequences in the reference metaphase chromosomes by prehybridization with appropriate blocking nucleic acids, and/or blocking those repetitive sequences in the labeled DNA by prehybridization with appropriate blocking nucleic acid sequences, and/or including such blocking nucleic acid sequences for said repetitive sequences during said hybridization;

e) rendering the bound, differently labeled DNA sequences visualizable, if necessary;

f) observing and/or measuring the intensities of the signals from each subject DNA, and the relative intensities, as a function of position along the reference metaphase chromosomes; and

g) comparing the relative intensities among different locations along the reference metaphase chromosomes wherein the greater the intensity of the signal at a location due to one subject DNA relative to the intensity of the signal due to the other subject DNA at that location, the greater the copy number of the sequence that binds at that location in the first subject cell or cell population relative to the copy

number of the substantially identical sequence in the second subject cell or cell population that binds at that location.

10. A method of quantitatively comparing copy numbers of different DNA sequences in one subject cell or cell population relative to copy numbers of substantially identical sequences in another subject cell or cell population, said method comprising steps (a) through (e) of Claim 9 and steps of:

f. measuring the intensities of the signals from each of the bound subject DNAs and calculating the ratio of the intensities as a function of position along the reference metaphase chromosomes to form a ratio profile; and

g. quantitatively comparing the ratio profile among different locations along the reference metaphase chromosomes, said ratio profile at each location being proportional to the ratio of the copy number of the DNA sequence that binds to that location in the first subject cell or cell population to the copy number of a substantially identical sequence in the second cell or cell population.

11. A method to determine the ratio of copy numbers of different DNA sequences in one subject cell or cell population to copy numbers of substantially identical sequences in another cell or cell population, said method comprising the steps of (a) through (f) of Claim 10 and steps of:

g. determining the average copy number of a calibration sequence in both subject cells or cell populations, said calibration sequence being substantially identical to a single copy sequence in the reference metaphase cells; and

h. normalizing the ratio profile calculated in (f) so that at the calibration position, the ratio profile is equal to the ratio of the average copy numbers determined in (g), the normalized ratio profile at any other location along the reference metaphase chromosomes thereby giving the ratio

of the copy numbers of the DNA sequences in the two subject DNAs that would bind at that location.

12. The method according to Claim 11 further comprising determining the ratio of copy numbers of DNA sequences in more than two subject DNAs wherein the comparing is done pairwise between signals from each subject DNA.

13. The method according to Claim 11 further comprising determining the copy numbers for more than one calibration position in step (g); and doing the normalizing in step (h) to obtain the best fit of the ratio profile to the calibration positions.

14. A method for comparing copy numbers of different DNA sequences in a test cell or cell population, said method comprising applying steps (a) through (f) of Claim 9 wherein one of the subject cells or cell populations is the test cell or cell population and the other is a normal cell or cell population, and step (g) of comparing the relative intensities among different locations along the reference metaphase chromosomes, wherein the greater the relative intensity at a location, the greater the copy number of the sequence in the test cell or cell population that binds to that location, except for sex chromosomes where the comparison needs to take into account the known differences in copy numbers of sequences in the sex chromosome in relation to those on the autosomes in the normal subject cell or cell population.

15. A method for comparing the copy number of different DNA sequences in a test cell or cell population, said method comprising applying steps (a) through (f) of Claim 10 wherein one of the subject cells or cell populations is the test cell or cell population, and the other is a standard cell or cell population wherein the known copy numbers of the DNA sequences that bind to different positions on the reference metaphase chromosomes is known, and the following steps:

g. adjusting the ratio profile at each location along the reference metaphase chromosomes by multiplying the ratio profile by the known copy number of DNA sequences in the standard cell or cell population that bind there; and

h. comparing the adjusted ratio profiles at different locations along the reference metaphase chromosomes wherein the greater the adjusted ratio profile at a location, the greater the copy number of the DNA sequence in the test cell or cell population that binds there.

16. A method of determining the ratios of the copy numbers of different DNA sequences in a test cell or cell population, said method comprising applying steps (a) through (g) of Claim 15 and calculating the ratio of the copy number of a DNA sequence in the test cell or cell population that binds to one location on the reference metaphase chromosomes to the copy number of a sequence that binds to another location by dividing the adjusted ratio profile at the location of the first sequence by that at the location of the second.

17. A method for determining the copy number of different DNA sequences in a test cell or cell population, said method comprising

applying steps (a) through (g) of Claim 15;

h. determining the copy number of a calibration sequence in the test cell or cell population that is substantially identical to a single copy sequence in the reference cells; and

i. normalizing the adjusted ratio profile so that at the location of the calibration sequence on the reference metaphase chromosomes, the normalized, adjusted ratio profile is equal to the copy number of the calibration sequence determined in (h), the value of the normalized, adjusted ratio profile at another location then being equal to the copy number of the DNA sequence in the test cell or cell population that binds at that location.

18. The method of Claim 17 wherein two or more calibration sequences are used, and the adjusted ratio profile is normalized to get the best fit to the copy numbers of the ensemble of calibration sequences.

19. A method of comparing copy numbers of different RNA sequences in one subject cell or cell population relative to copy numbers of substantially identical sequences in another cell or cell population, said method comprising the steps of:

a) extracting the RNA from both of the subject cells or cell populations;

b) amplifying said extracted subject RNAs, if necessary;

c) labeling each of the subject RNAs differently;

d) hybridizing said differentially labeled subject RNAs <u>in situ</u> to reference metaphase chromosomes after substantially removing from the labeled RNAs those repetitive sequences that could bind to multiple loci in the reference metaphase chromosomes, and/or after blocking the binding sites for those repetitive sequences in the reference metaphase chromosomes by prehybridization with appropriate blocking nucleic acids, and/or blocking those repetitive sequences in the labeled RNA by prehybridization with appropriate blocking nucleic acid sequences, and/or including such blocking nucleic acid sequences for said repetitive sequences during said hybridization;

e) rendering the bound, differently labeled RNA sequences visualizable, if necessary;

f) observing and/or measuring the intensities of the signals from each subject RNA, and the relative intensities as a function of position along the reference metaphase chromosomes; and

g) comparing the relative intensities among different locations along the reference metaphase chromosomes wherein the greater the intensity of the signal at a location due to one subject RNA relative to the intensity of the signal due to the other subject RNA at that location, the greater the

copy number of the sequence that binds at that location in the first subject cell a cell population relative to the copy number of the substantially identical sequence in the second subject cell or cell population.

20. A method of detecting amplification of a certain sequence or group of sequences in a subject cell or cell population, comprising steps (a) through (e) of Claim 1 wherein the <u>in situ</u> hybridization is targeted to antenna cells in which the DNA sequence(s) to be tested for is or are amplified, and examining the reference cell for regions that are hybridized significantly more intensely than others, the presence of such regions indicating amplifications of the sequence(s) which are being tested.

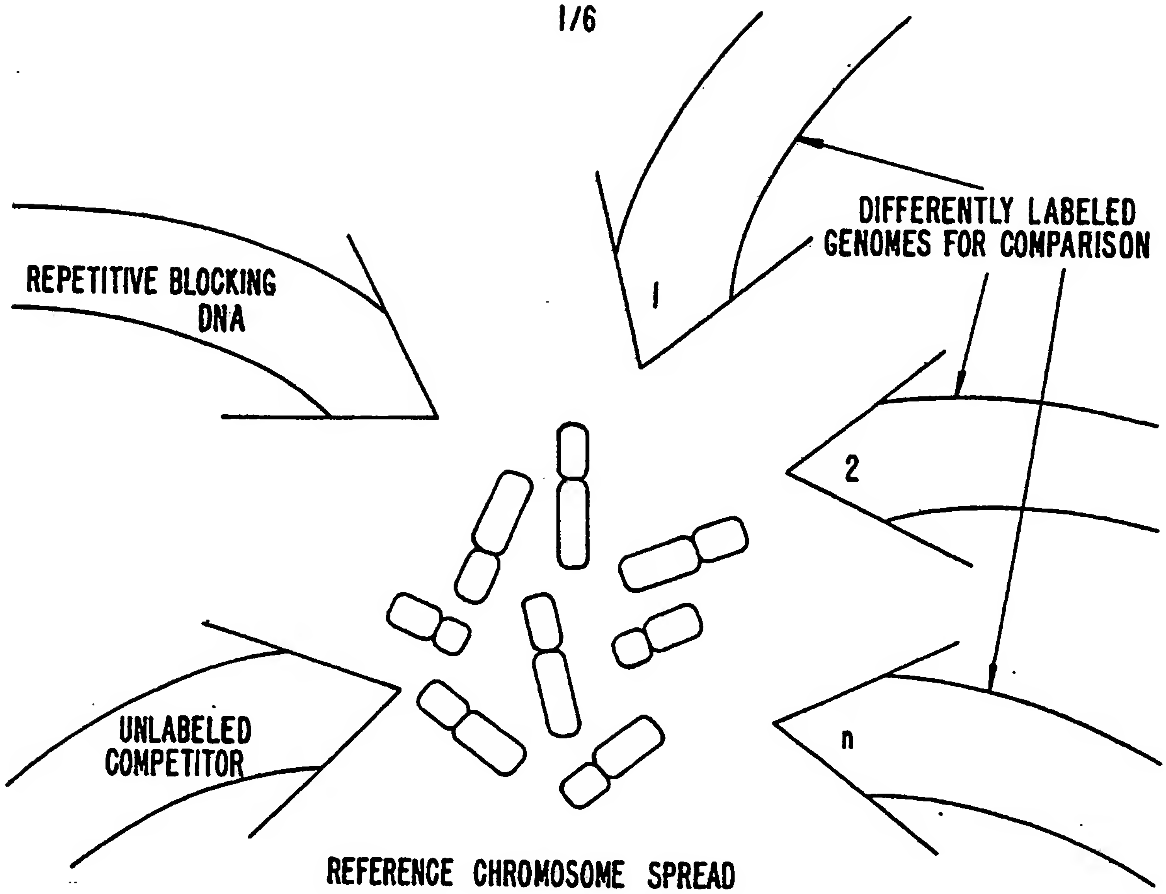

DIFFERENTLY LABELED
GENOMES FOR COMPARISON
REPETITIVE BLOCKING
DNA
1
2
n
UNLABELED
COMPETITOR
REFERENCE CHROMOSOME SPREAD

FIG. 1.

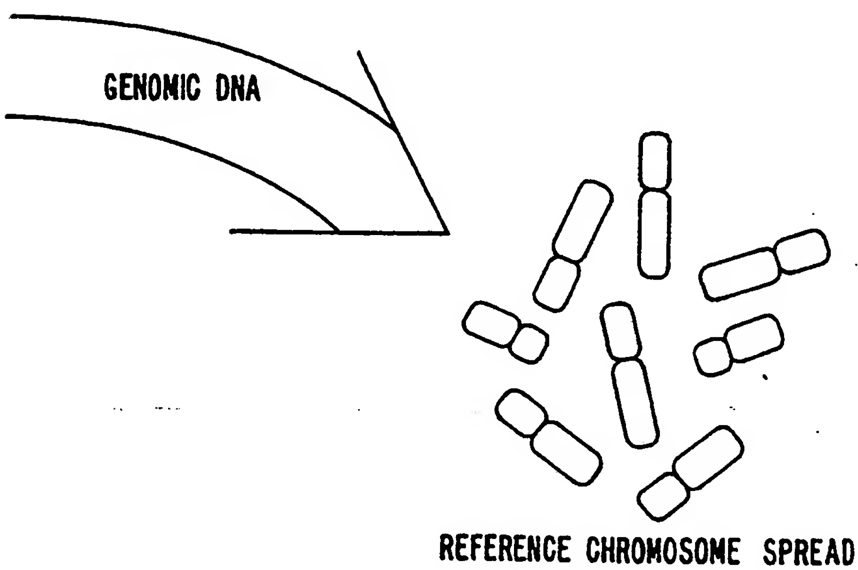
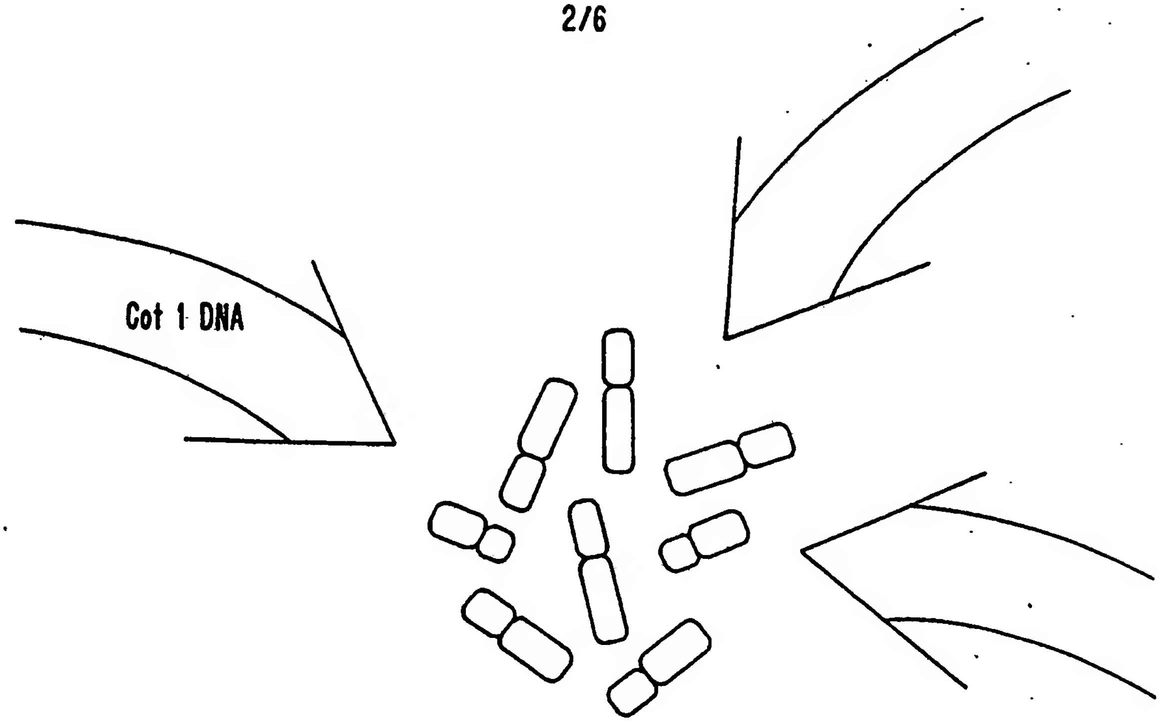
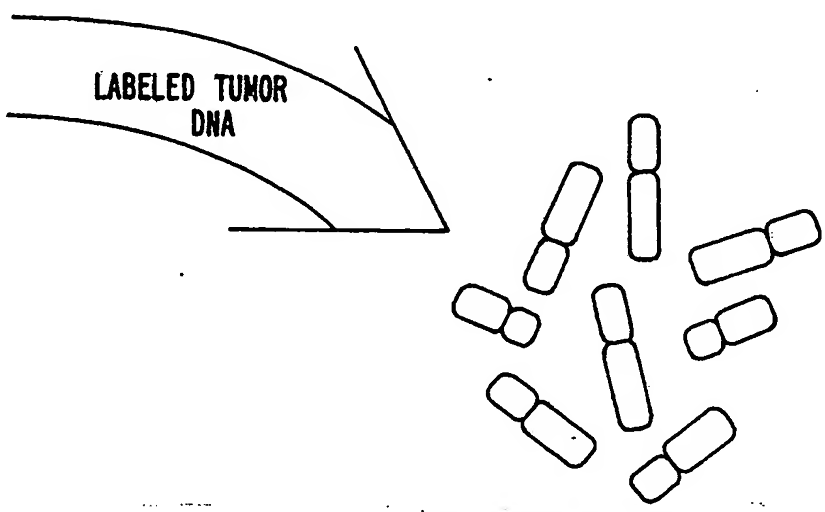

GENOMIC DNA
REFERENCE CHROMOSOME SPREAD

FIG. 2A.



REFERENCE CHROMOSOME SPREAD

FIG. 2B.



HUMAN CHROMOSOME SPREAD

FIG. 3A.

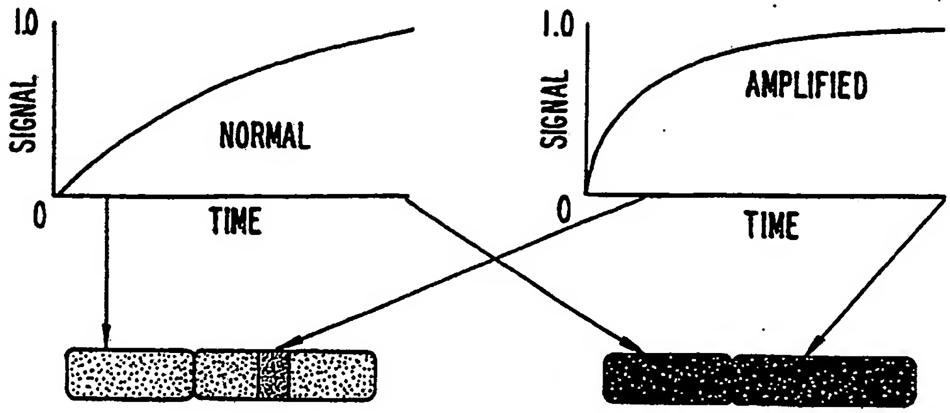


FIG. 3B.

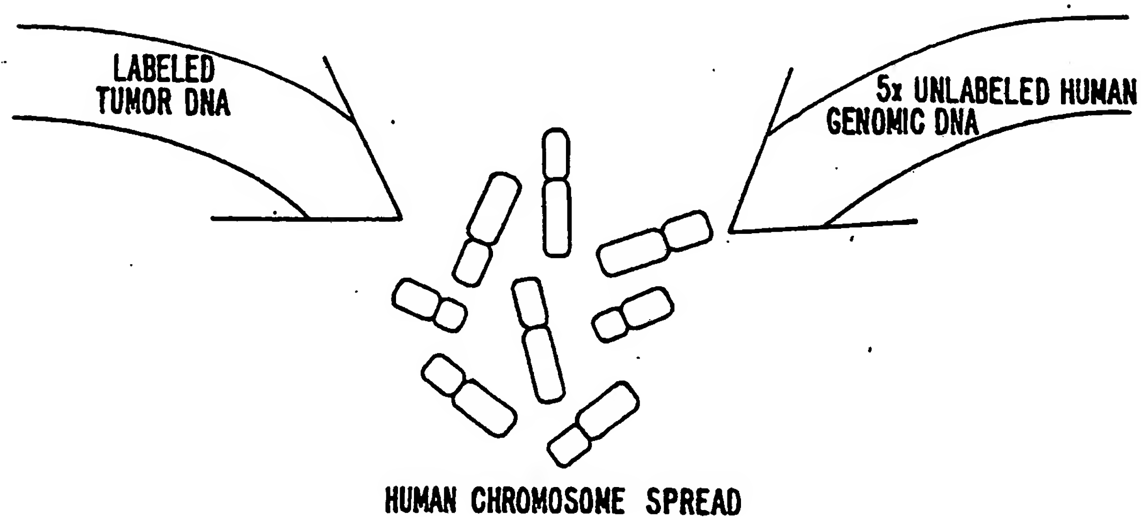


FIG. 4A.

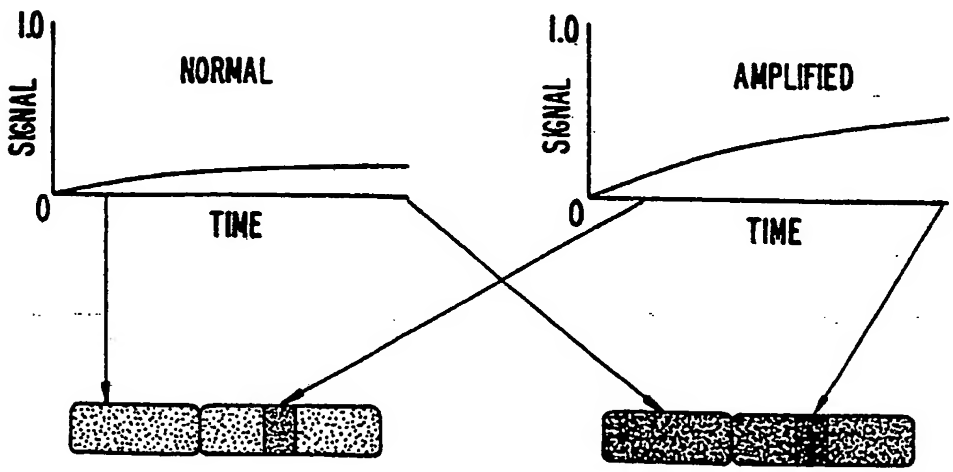


FIG. 4B.

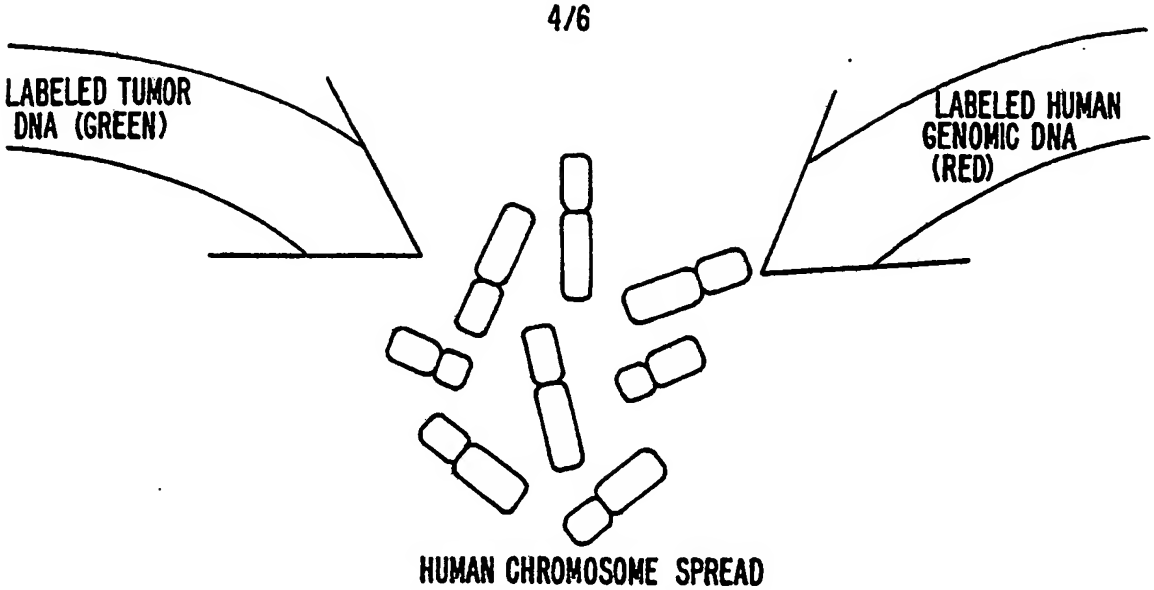

LABELED TUMOR DNA (GREEN)
LABELED HUMAN GENOMIC DNA (RED)
HUMAN CHROMOSOME SPREAD

FIG. 5.

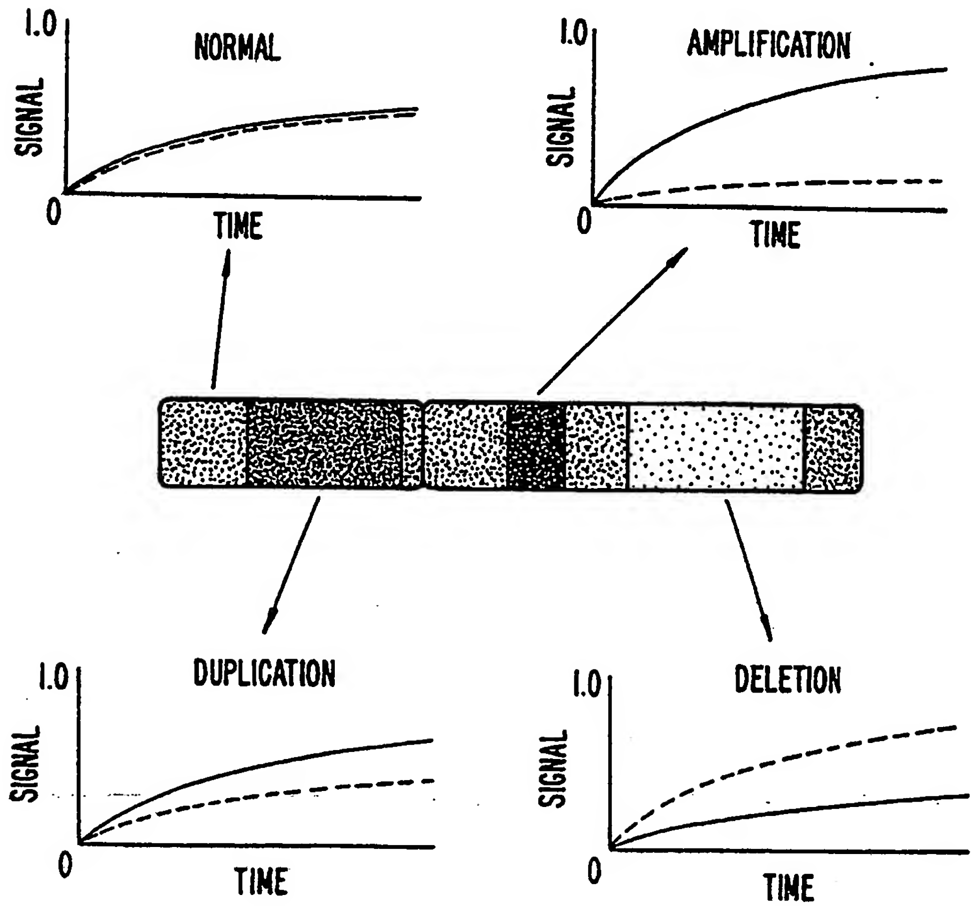

1.0
NORMAL
SIGNAL
0
TIME
1.0
AMPLIFICATION
SIGNAL
0
TIME
1.0
DUPLICATION
SIGNAL
0
TIME
1.0
DELETION
SIGNAL
0
TIME

FIG. 6.

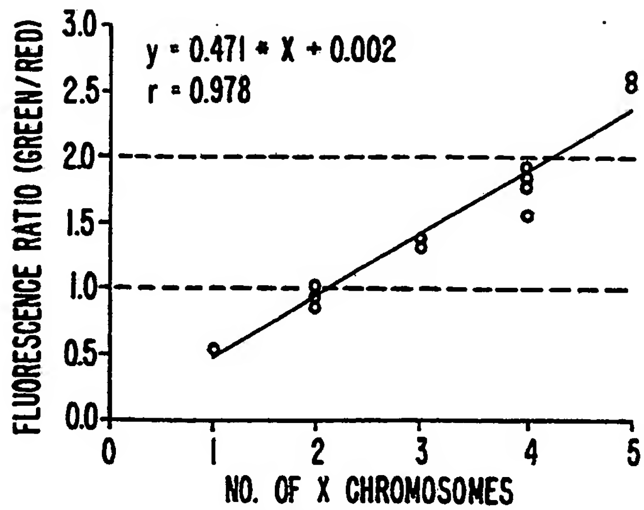


FIG. 7.

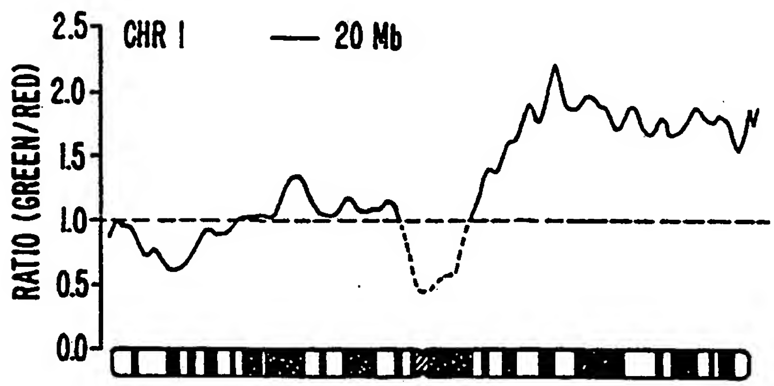


FIG. 8A.

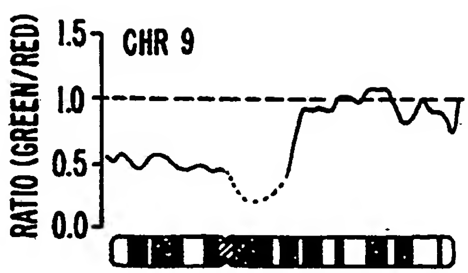


FIG. 8B.

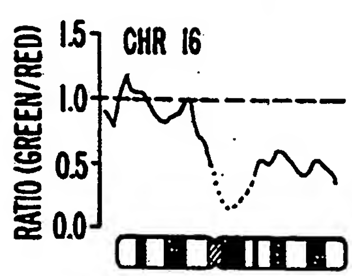


FIG. 8D.

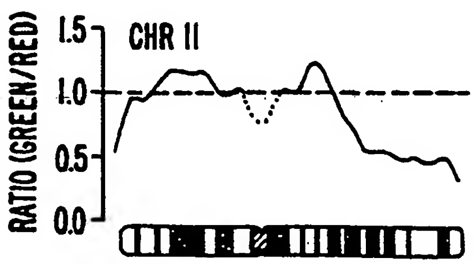


FIG. 8C.

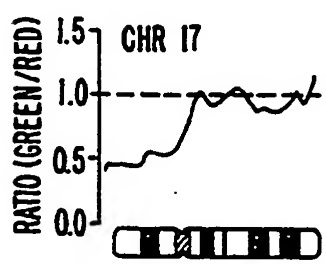


FIG. 8E.

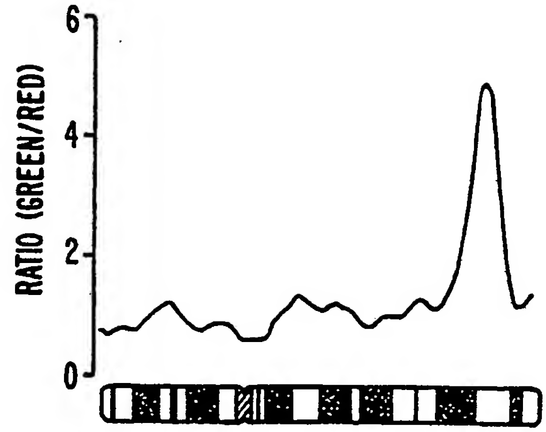


FIG. 9A.

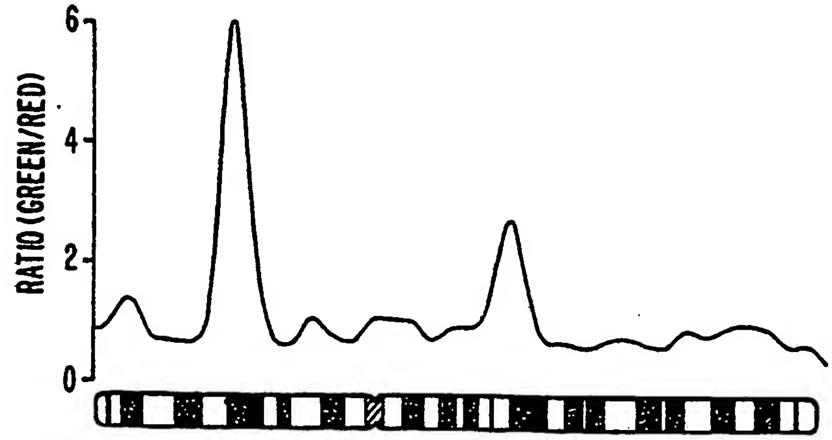


FIG. 9B.

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 93/01775

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12Q1/68

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA. vol. 83, May 1986, WASHINGTON US pages 2934 - 2938 D. PINKEL ET AL. cited in the application see the whole document ---	1
A	CHROMOSOMA vol. 83, no. 2, 1981, BERLIN,GERMANY pages 159 - 168 H.W. BROCK ET AL. see the whole document, especially the abstract --- -/--	1

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
29 APRIL 1993	18. 05. 93
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	LUZZATTO E.R.

3

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	WO,A,9 005 789 (YALE UNIVERSITY) 31 May 1990 cited in the application see page 3, line 1 - page 4, line 3 see page 12, line 14 - page 13, line 19; claims; figure 1 ---	1,8,9
A	EP,A,0 430 402 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 5 June 1991 cited in the application see page 6, line 54 - page 10, line 6; claims ---	1
P,X	SCIENCE vol. 258, 30 October 1992, LANCASTER, PA US pages 818 - 821 A. KALLIONIEMI ET AL. see the whole document -----	1,10,11

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

US 9301775
SA 71380

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on 29/04/93
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9005789	31-05-90	AU-A- 4642089	12-06-90
		CA-A- 2003051	15-05-90
		CN-A- 1046392	24-10-90
		EP-A- 0444115	04-09-91
		JP-T- 4502855	28-05-92
EP-A-0430402	05-06-91	AU-A- 5898790	06-06-91
		CA-A- 2021489	20-01-91
		JP-A- 3224499	03-10-91

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

negative selection is increased. This results in an A2-restricted repertoire that contains receptors with higher than usual affinity for A2 per se. When provided with the A2/K^b molecule, murine CD8 can participate more fully in interaction with the restriction element, thus elevating the avidity of the interaction above the threshold required for the induction of detectable lytic activity in the absence of antigen. However, because CD8 is also required for T cell stimulation, our results are also consistent with a model in which only the affinity cutoff for negative selection is increased and the observed recognition of A2 per se is due to an increase in the affinity requirement for antigen responsiveness rather than positive selection.

REFERENCES AND NOTES

1. M. M. Davis and P. J. Bjorkman, *Nature* 334, 395 (1988).
2. R. D. Salter *et al.*, *ibid.* 345, 41 (1990).
3. J. M. Connolly, T. A. Potter, E.-M. Wormstall, T. H. Hansen, *J. Exp. Med.* 168, 325 (1988); T. A. Potter, T. V. Rajan, R. F. Dick II, J. A. Bluestone, *Nature* 337, 73 (1989).
4. M. J. Bevan, *Nature* 269, 417 (1977); R. M. Zinkernagel *et al.*, *J. Exp. Med.* 147, 882 (1978); W. C. Sha *et al.*, *Nature* 335, 271 (1988); H. S. Teh *et al.*, *ibid.*, p. 229; P. Kisielow, H. S. Teh, H. Bluthmann, H. von Boehmer, *ibid.*, p. 730; L. J. Berg *et al.*, *Cell* 58, 1035 (1989); J. Kaye *et al.*, *Nature* 341, 746 (1989); W. C. Sha *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 87, 6186 (1990).
5. J. W. Kappler, N. Roehm, P. Marrack, *Cell* 49, 273 (1987); J. W. Kappler, U. Staerz, J. White, P. Marrack, *Nature* 332, 35 (1988); H. R. MacDonald *et al.*, *ibid.*, p. 40; W. C. Sha *et al.*, *ibid.* 336, 73; H. S. Teh, H. Kishi, B. Scott, H. von Boehmer, *J. Exp. Med.* 169, 795 (1989).
6. N. A. Lee, D. Y. Loh, E. Lacy, *J. Exp. Med.* 175, 1013 (1992); M. Blum *et al.*, *Cell* 69, 1097 (1992).
7. R. C. Duke, *J. Exp. Med.* 170, 59 (1989).
8. M. J. Irwin, W. R. Heath, L. A. Sherman, *ibid.*, p. 1091.
9. N. L. Samberg, E. C. Scarlett, H. J. Strauss, *Eur. J. Immunol.* 19, 2349 (1989).
10. U. Kalinke, B. Arnold, G. J. Hammerling, *Nature* 348, 642 (1990).
11. A. Vitiello, D. Marchesini, J. Furze, L. A. Sherman, R. W. Chestnut, *J. Exp. Med.* 173, 1007 (1991).
12. F. H. Gotch *et al.*, *Nature* 326, 881 (1987).
13. V. H. Engelhard, E. Lacy, J. P. Ridge, *J. Immunol.* 146, 1226 (1991).
14. D. La Face *et al.*, *FASEB J.* 6 (Abstr. 4474), A1709 (1992).
15. A. L. Ingold, C. Landel, C. Knall, G. A. Evans, T. A. Potter, *Nature* 352, 721 (1991); C. J. Aldrich *et al.*, *ibid.*, p. 718.
16. B. H. Koller and H. T. Orr, *J. Immunol.* 134, 2727 (1985).
17. A. Vitiello and L. A. Sherman, *ibid.* 131, 1635 (1983).

18. We thank J. Price for assistance in production of transgenic mice and C. Wood for excellent secretarial assistance. Supported by grant CA 25803 from the National Cancer Institute of the National Institutes of Health.

9 July 1992; accepted 23 September 1992

Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors

Anne Kallioniemi,* Olli-P. Kallioniemi, Damir Sudar, Denis Rutovitz, Joe W. Gray, Fred Waldman, Dan Pinkel

Comparative genomic hybridization produces a map of DNA sequence copy number as a function of chromosomal location throughout the entire genome. Differentially labeled test DNA and normal reference DNA are hybridized simultaneously to normal chromosome spreads. The hybridization is detected with two different fluorochromes. Regions of gain or loss of DNA sequences, such as deletions, duplications, or amplifications, are seen as changes in the ratio of the intensities of the two fluorochromes along the target chromosomes. Analysis of tumor cell lines and primary bladder tumors identified 16 different regions of amplification, many in loci not previously known to be amplified.

The discovery of genetic changes involved in the development of solid tumors has proven difficult. Karyotyping is impeded by the low number of high-quality metaphase spreads and the complex nature of chromosomal changes (*1*). Molecular genetic studies of isolated tumor DNA have been more successful and have been used to detect common regions of allelic loss, mutation, or amplification (*2*, *3*). However, such molecular methods are highly focused; they target one specific gene or chromosome region at a time and leave the majority of the genome unexamined.

We have developed a molecular cytogenetic method, comparative genomic hybridization (CGH), that is capable of detecting and mapping relative DNA sequence copy number between genomes. A copy number karyotype can be generated for a tumor by the comparison of DNAs from malignant and normal cells, thereby identifying regions of gain or loss of DNA. In this application of CGH, biotinylated total tumor DNA and digoxigenin-labeled normal genomic reference DNA are simultaneously hybridized to normal metaphase spreads in the presence of unlabeled Cot-1 blocking DNA (*4*). Hybridization of tumor DNA is detected with green-fluorescing fluorescein isothiocyanate (FITC)–avidin, and the reference DNA hybridization is detected with red-fluorescing rhodamine antidigoxigenin (*5*). The relative amounts of tumor and reference DNA bound at a given chromosomal locus are dependent on the relative abundance of those sequences in the two DNA samples and can be quantitated by measurement of the ratio of green-to-red fluorescence. The reference DNA serves as a control for local variations in the ability to hybridize to target chromosomes. Thus, gene amplification or chromosomal duplication in the tumor DNA produces an elevated green-to-red ratio, and deletions or chromosomal loss cause a reduced ratio. The Cot-1 DNA included in the hybridization inhibits binding of the labeled DNA to the centromeric and heterochromatic regions, so these regions are excluded from the analysis. The fluorescence signals are quantitatively analyzed by means of a digital image analysis system (*6*). A software program integrates the green and red fluorescence intensities in strips orthogonal to the chromosomal axis, subtracts local background, and calculates intensity profiles for both colors and the green-to-red ratio along each chromosome.

The ability of CGH to quantitate changes in sequence copy number that affect an entire chromosome was tested with five fibroblast cell lines having one to five copies of the X chromosome and two copies of each autosome (*7*). Hybridization of DNA from a 45,X0 cell line (in green), together with normal female reference DNA (in red) to a normal male metaphase spread, resulted in a uniform green-red staining of the autosomes, whereas the X chromosome appeared more red (Fig. 1A). Hybridizations with DNA from cell lines carrying two, three, four, or five copies of the X chromosome resulted in increasingly strong green fluorescence from the X chromosome relative to the autosomes. The average green-to-red fluorescence ratio of the X chromosome (Fig. 1B), when normalized to the average ratio for the autosomes within the same metaphase spread, increased linearly with increasing number of X chromosomes [correlation coefficient (r) = 0.978]. Thus, CGH can quantitatively distinguish a change of plus or minus one copy of a chromosome at least up to four copies.

We used CGH to generate a complete copy number karyotype for a near-diploid breast cancer cell line, 600PE. According to the published karyotype (*8*), 600PE is near-diploid with five marker chromo-

A. Kallioniemi, O.-P. Kallioniemi, D. Sudar, J. W. Gray, F. Waldman, D. Pinkel, Division of Molecular Cytometry, Department of Laboratory Medicine, University of California at San Francisco, San Francisco, CA 94143. D. Rutovitz, Medical Research Council Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, United Kingdom.

*To whom correspondence should be addressed.

somes. The CGH with biotinylated 600PE DNA (in green) and normal digoxigenin-labeled reference DNA (in red) revealed the following relative copy number changes: gain of 1q and loss of 9p, 16q, 17p, and distal 11q. The green-to-red ratio profiles for these aberrant chromosomes are shown in Fig. 2. Fluorescence in situ hybridization (FISH) with locus-specific 16p and 16q cosmid probes to 600PE cells (9) found two signals per nucleus for 16p and one for 16q that permitted calibration of a green-to-red ratio of 1.0 as indicating two copies of a sequence.

Thus, if the absolute copy number of any region in a tumor genome is known, relative numbers can be converted to actual copy numbers at all loci. The CGH results differed from the originally published karyotype in the region of 16p and proximal 1p. This discrepancy was subsequently resolved based on the FISH analysis with locus-specific probes that indicated that the components of one of the marker chromosomes had been misinterpreted by conventional cytogenetic analysis (8). CGH with DNA from two fibroblast cell lines (*10*) detected small interstitial deletions around the *RB1* locus: del (13) (pter>q14.1::q21.2>qter) and del (13) (pter>q14.1::q22.1>qter). On the basis of the CGH analysis and measurement of the deletion size as a fraction of the length of chromosome 13 (total length 111 Mb), these deletions were estimated to span about 10 and 20 Mb, respectively. Thus, it is possible that CGH could be used to screen DNA samples from solid tumors in order to identify physical deletions that may uncover recessive mutant tumor suppressor genes.

The CGH technique was evaluated for its ability to detect increased gene copy number with cell lines that contained previously reported amplification of oncogenes. CGH was performed with DNA from a colon cancer cell line, COLO 320HSR (Fig. 3A), known to contain more than a 50-fold amplification of a 300-kb region around the *myc* oncogene (*11*). The expected high green-to-red ratio at 8q24 that corresponds to the *myc* locus is clear. The height of the peak does not quantitatively reflect the level of amplification because the fluorescent signal spreads over a region of the chromosome that is larger than the length of the amplicon. This is apparently a result of the complex organization of the target DNA in the denatured

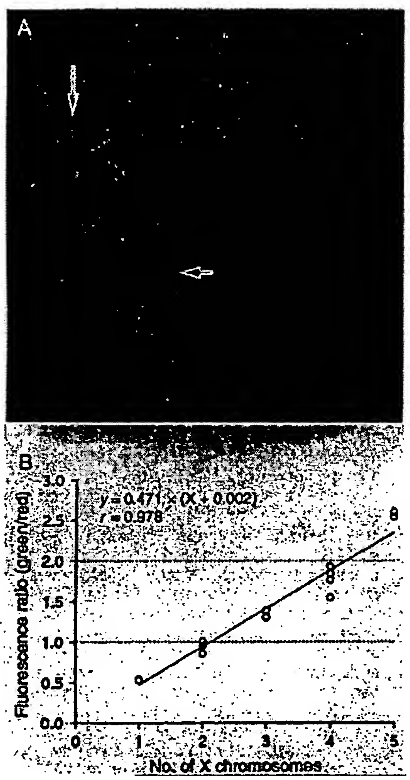


Fig. 1. (**A**) Hybridization of DNA from the 45,XO cell line (green) and normal female reference DNA (red) to a normal male metaphase spread. The reddish color of the X chromosome (large arrow) as compared with autosomes reflects the lower relative copy number of the X chromosome sequences in this cell line. Faint staining of a small part of the Y chromosome (small arrow) is a result of the homologous sequences in the pseudo-autosomal region. (**B**) Correlation of the number of X chromosomes in five fibroblast cell lines and the average green-to-red ratio of the X chromosome relative to the same ratio for the autosomes.

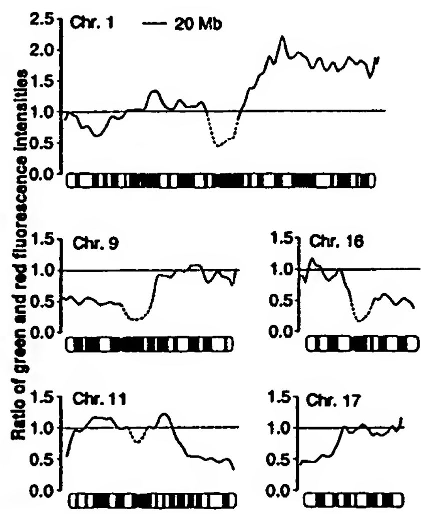


Fig. 2. Green-to-red fluorescence ratio profiles of chromosomes 1, 9, 11, 16, and 17 after hybridization with 600PE DNA (green) and normal reference DNA (red). Profiles reflect the relative copy number of the chromosomal regions. FISH of 600PE interphase and metaphase cells indicated that the copy number for 16p loci was 2. Thus, the ratio of 1.0 indicates two copies of the sequence throughout the genome. The dip in the profile at 1p34 through 1p36 may represent a previously unsuspected small interstitial deletion but has not been verified independently with specific probes for this region. Centromeric and heterochromatic regions of the genome are not included in the analysis because the Cot-1 DNA partially blocks signals in these regions, and the large copy number polymorphisms between individuals at these loci make the ratio data unreliable. Chr., chromosome.

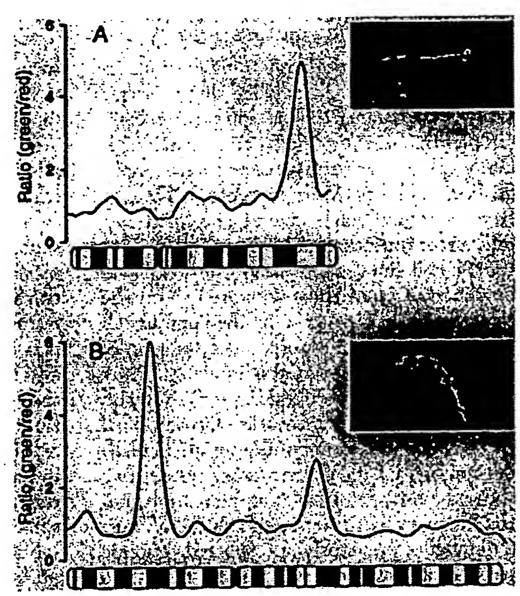


Fig. 3. Green-to-red fluorescence ratio profiles of chromosome 8 (**A**) and chromosome 2 (**B**) after hybridization with COLO 320HSR and NCI-H69 cell line DNAs, respectively (green). Normal reference DNA included in the hybridization is shown in red. The inserts show the overlaid green and red fluorescence images of the chromosomes and the chromosomal medial axis drawn by the image analysis program. In (A), the *myc* locus at 8q24 shows a highly elevated green-to-red ratio, which is compatible with the known high-level amplification of *myc* in the COLO 320HSR cell line. In (B), three regions of amplification are seen on chromosome 2. The signal at 2p24 corresponds to the location of *N-myc* known to be amplified in the NCI-H69 cell line. The two other regions with a highly increased fluorescence ratio, at 2p21 and 2q21, were not known to be amplified.

Table 1. Mapping of amplified sequences in established cancer cell lines and primary tumors by CGH. Cytogenetic information is based on the American Type Culture Collection catalog of cell lines and hybridomas (1992). DM, double minute chromosomes; HSR, homogeneously staining regions.

Specimen	Origin	Amplification by CGH*	Cytogenetic evidence of gene amplification
		Cell lines	
5637	Bladder	3p25, 6p22	DM
SK-BR-3	Breast	8q24 (*myc*), 8q21, 17q12 (*erb*B2), 20q13	—
COLO 205	Colorectal	6p21, 6q24	—
NCI-H508	Colorectal	14q12-q13	DM
SW480	Colorectal	8q24 (*myc*)	DM
SW620	Colorectal	16q21-q23	HSR
WiDr	Colorectal	8q23-q24 (*myc*)	—
SK-N-MC	Neuroblastoma	8q24 (*myc*)	DM
Calu-3	Small cell lung	8p12-p21, 8qtel, 17q12 (*erb*B2)	HSR
Calu-6	Small cell lung	13q32-q34	—
NCI-H69	Small cell lung	2p24 (N-*myc*), 2p21, 2q21	—
		Primary tumors	
UR140	Bladder carcinoma	16q21-q22	—
UR145	Bladder carcinoma	6p22	—

*The oncogene most likely involved in this amplification is shown in parentheses.

chromosomes. The eightfold amplification of the *erbB2* oncogene in the SK-BR-3 breast cancer cell line also was detectable as a hybridization signal at 17q12 (Table 1) with CGH. High-level amplifications such as these could also be detected in single-color hybridizations with the use of only labeled tumor DNA.

Cytogenetic and molecular studies of primary tumors and cell lines often reveal homogeneously staining regions and double minute chromosomes that do not involve known oncogenes (*12*). The CGH technique provides a means to detect and map such sequences. Table 1 contains a summary of the analysis with CGH of 11 cancer cell lines. Data in the table are based on the visual inspection of a large number of metaphase spreads and on detailed digital image analysis of four to six metaphases for each sample. Sixteen amplified loci were mapped, many at regions of the genome where amplification had not been suspected. Thus, a large variety of genes may be amplified during cancer initiation and progression. In 5 of the 11 cell lines, more than one locus was amplified. Two or three separate loci on the same chromosome were amplified in four cell lines, which suggests a spatial clustering of chromosomal locations that undergo DNA amplification (Table 1 and Fig. 3B). We also applied CGH to identify and map amplified DNA sequences in uncultured primary bladder tumors. Of the seven tumors tested, two showed evidence of DNA amplification, but the loci were not the same (Table 1). Thus, a number of previously unsuspected genomic regions that might contain genes important for cancer progression have been identified. Further studies will be required to determine which loci contain novel oncogenes and which represent coincidental, random DNA amplification characteristic of general genomic instability.

The detection and mapping of unknown amplified sequences that typically span several hundred kilobases to a few megabases demonstrates the usefulness of the CGH technique for rapid identification of regions of the genome that may contain oncogenes. Similarly, detection of deletions may facilitate identification of regions that contain tumor suppressor genes. The ability to survey the whole genome in a single hybridization is a distinct advantage over allelic loss studies by restriction fragment length polymorphism (RFLP) that target only one locus at a time. RFLP is also restricted by the availability and informativeness of polymorphic probes. However, further studies are required to establish to what extent allelic losses in tumors are caused by physical deletions. In clinical specimens, the detection of small copy number differences, such as those associated with deletions, is more difficult than in cell lines because of the admixture of DNA from contaminating normal cells and because of intratumor genetic heterogeneity. Like RFLP, CGH also emphasizes the detection of aberrations that are homogeneous in the cell population and averages those that are heterogeneous. Currently, sensitivity is primarily limited by the granularity of the hybridization signals in the metaphase chromosomes. We expect further improvements in sensitivity will be achieved by optimization of the probe concentration and labeling and by the averaging of the green-to-red fluorescence ratios from several metaphase spreads. CGH comparisons other than those described here—for example, between primary tumors and their metastases—should also be informative.

REFERENCES AND NOTES

1. J. R. Teyssier, *Cancer Genet. Cytogenet.* 37, 103 (1989).
2. E. R. Fearon and B. Vogelstein, *Cell* 61, 759 (1990); T. Sato *et al.*, *Cancer Res.* 50, 7184 (1990).
3. K. Alitalo and M. Schwab, *Adv. Cancer Res.* 47, 235 (1986); M. Schwab and L. C. Amler, *Genes Chrom. Cancer* 1, 181 (1990).
4. DNAs were labeled with biotin-14–deoxyadenosine triphosphate or digoxigenin-11–deoxyuridine triphosphate by nick translation. The optimal size for double-stranded probe fragments was 600 to 1000 bp. Lymphocyte metaphase preparations were denatured in 70% formamide and 2× SSC [1× SSC is 0.15 M NaCl and 0.015 M Na citrate (pH 7)] at 70°C for 2 to 3 min and dehydrated in a sequence of 70%, 85%, and 100% ethanol. The slides were air-dried, treated with Proteinase K (0.2 μg/ml) (Boehringer Mannheim) for 7.5 min at 37°C in 20 mM tris and 2 mM $CaCl_2$ (pH 7.5), and dehydrated by ethanol as above. Sixty nanograms of biotinylated test DNA, 60 ng of digoxigenin-labeled normal reference DNA, and 5 μg of unlabeled Cot-1 DNA (BRL, Gaithersburg, MD) were precipitated with ethanol and dissolved in 10 μl of 50% formamide, 10% dextran sulfate, and 2× SSC (pH 7). This probe mixture was denatured at 70°C for 5 min, allowed to reanneal at 37°C for 60 min, and hybridized to normal male metaphase chromosomes for 3 to 4 days at 37°C.
5. Slides were washed at 45°C three times in 50% formamide–2× SSC (pH 7), twice in 2× SSC, and once in 0.1× SSC (10 min for each wash). After washing, the slides were immunocytochemically stained at room temperature in three 30-min steps: (i) FITC-avidin (5 μg/ml; Vector Laboratories, Burlingame, CA) and antidigoxigenin-rhodamine (2 μg/ml; Boehringer Mannheim); (ii) antiavidin (5 μg/ml; Vector Laboratories); and (iii) FITC-avidin (5 μg/ml). Nuclei were counter-stained with 0.8 μM 4,6-diamino-2-phenylindole (DAPI) in an antifade solution [G. D. Johnson and J. G. M. Nogueria, *J. Immunol. Methods* 43, 349 (1981)]. A Zeiss fluorescence microscope equipped with a double bandpass filter (Chroma Technology, Brattleboro, VT) was used for simultaneous visualization of FITC and rhodamine signals.
6. The Quantitative Image Processing System (QUIPS) is based on a Nikon Microphot SA fluorescence microscope. A cooled charge-coupled device camera (Photometrics, Tucson, AZ) interfaced to a SUN 4/330 work station was used for image acquisition. Programs for the acquisition and display of multicolor images were developed at our laboratory. The chromosomes were identified, and the band locations of the hybridization signals were assigned on the basis of a digital contrast-enhanced image of the DAPI counterstain. Fluorescence ratio profiles along the chromosomes are extracted by means of "WOOLZ" software package (developed at the Medical Research Council, Edinburgh) as follows. The DAPI image is used to set the morphological boundary of each chromosome by thresholding. The chromosome outline is smoothed by an *n* number of opening and closing operations, and the medial axis of the chromosome is determined. The DAPI image is expanded outward in all directions until the intensity field levels off (when background is reached) or begins to rise (because of an adjacent chromosome). The intensity profile of each image along the medial axis and within the expanded DAPI image is then calculated by a summation of the green and red fluorescence pixel values along the sequence of lines perpendicular to and spaced at unit distance along the medial axis. Modal green and red intensity values

that correspond to the expanded DAPI image are used to represent background fluorescence and are used as the intensity origin.
7. GMO1723 (45,XO), GM08399 (46,XX), GMO4626 (47,XXX), GMO1415E (48,XXXX), and GMO5009B (49,XXXXX) from the National Institute of General Medical Sciences (NIGMS) repository (Camden, NJ).
8. H. S. Smith *et al.*, *J. Natl. Cancer Inst.* 78, 611 (1987). According to the published karyotype, 600PE contains only one normal copy of chromosomes 1, 9, 11, 13; 16, and 17 and has five marker chromosomes: t(1q:13q), 1p-(p22), inv(11)(p15q13), t(9q:17q), and inv(1)(p36q21). FISH of 600PE interphase nuclei and metaphase chromosomes (*9*) showed two signals with 16p cosmid probes and one signal with 16q cosmids. The other 16p signal was located on the second marker chromosome, which indicates that this marker had been misclassified by conventional cytogenetic analysis and was actually a t(1q:16p).
9. M. Sakamoto, personal communication.
10. GMO5877 and GMO1142A come from the NIGMS repository (Camden, NJ).
11. K. W. Kinzker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 83, 1031 (1986).
12. C. Saint-Ruf *et al.*, *Genes Chrom. Cancer* 2, 18 (1990); S. Bruderlein, K. van der Bosch, P. Schlag, M. Schwab, *ibid.*, p. 63.

13. We thank H. S. Smith and L. Chen for supplying 600PE DNA, M. Sakamoto for mapping the chromosome 16 deletion in 600PE cells, P. Kyrölä for technical assistance, and the Los Alamos National Laboratory for chromosome 16 cosmids. This work was funded by the Office of Health and Environmental Research, Department of Energy, under contract DE-AC-03-76SF00098, and NIH grants CA 45919, CA 44768, and CA 47537 as well as by Imagenetics Inc. (Naperville, IL). A.K. and O.-P.K. have received support from the Finnish Science Academy, Finnish and Pirkanmaa Cancer Societies, Lahtikari Foundation, and Pirkanmaa Cultural Foundation (Finland).

8 June 1992; accepted 15 September 1992

TECHNICAL COMMENTS

Mantle Plumes and Mantle Sources

Basalts from many ocean islands define elongate arrays in Sr-Nd-Pb isotopic space; these likely reflect the dominance of binary mixing of mantle sources in intraplate volcanism. S. R. Hart *et al.* (*1*) observe that when these arrays are projected onto a ternary diagram bounded by mantle end-members DMM [depleted mid-ocean ridge basalt (MORB) mantle], HIMU (high U/Pb mantle), and EM1 (enriched mantle 1), they are subparallel and point toward a composition on the DMM-HIMU join. Hart *et al.* suggest this composition is associated with a high $^3He/^4He$ ratio (>30 R_A) and therefore cannot be a ubiquitous upper-mantle mixture of DMM and HIMU (both of which have $^3He/^4He$ < 9 R_A). They instead argue in favor of a new isotopic component, resident in a deep mantle "Focus Zone" (FOZO), which is characterized by a high $^3He/^4He$ ratio or which acquires helium with a high $^3He/^4He$ signature from the core.

There is an alternative interpretation of the He data. On the basis of binary isotope diagrams involving He, we (*2*) have suggested that the highest $^3He/^4He$ ratios are associated with intermediate $^{87}Sr/^{86}Sr$ and $^{143}Nd/^{144}Nd$ rather than with the highly depleted FOZO values. Mantle He-Sr-Nd-Pb systematics are consistent with five compositional endmembers—our finding of high $^3He/^4He$ Primitive Helium Mantle (PHEM) (*2*) and DMM-HIMU-EM1-EM2.

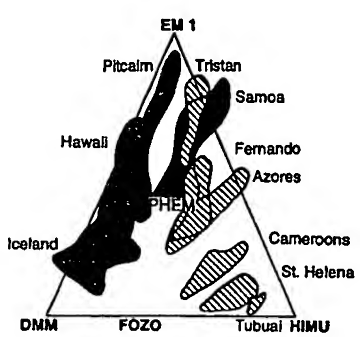


Fig. 1. Helium isotope affinity of some ocean island localities trending toward FOZO. Helium data are from standard sources [referred to in (*2*, *3*)] and our unpublished results. Striped areas indicate low ^3He; solid areas indicate high ^3He.

Combined He-Sr-Nd-Pb data are lacking for most ocean islands. However, many studies have demonstrated that ocean islands may be either "high ^3He" (all measured $^3He/^4He$ ratios ≥ MORB values) or "low ^3He" ($^3He/^4He$ ≤ MORB) (*3*). We know of no single ocean island that has $^3He/^4He$ ratios both greater and less than MORB (this also holds for most hot spot chains). Thus we can easily identify ocean islands that are tapping high $^3He/^4He$ material; if FOZO is characterized by a high $^3He/^4He$ ratio, then arrays trending toward it should be from high $^3He/^4He$ islands. We have replotted the mantle ternary (*1*) in Fig. 1 to indicate He affinities and the projection of the PHEM component into DMM-EM1-HIMU space. Many of the arrays tending toward FOZO are from low $^3He/^4He$ localities (Fig. 1), indicating that FOZO may not have a high $^3He/^4He$ ratio. We propose that some arrays mix to PHEM, and others to FOZO; isotopic relationships in DMM-HIMU-EM1-EM2-PHEM space are difficult to visualize in a ternary projection.

If FOZO does not have a high $^3He/^4He$ ratio, the need for a core or lower mantle helium source for FOZO is eliminated, and the component may be a DMM-HIMU mixture residing in the upper mantle. This has important implications for the origin and entrainment dynamics of mantle plumes. Additional studies of ocean islands, in which He is integrated with Sr-Nd-Pb, are required to accurately interpret $^3He/^4He$ ratios within the framework of the Sr-Nd-Pb mantle end-members.

K. A. Farley
Lamont-Doherty Geological Observatory, Palisades, NY 10964
H. Craig
Scripps Institution of Oceanography, La Jolla CA 92093

REFERENCES

1. S. R. Hart, E. H. Hauri, L. A. Oschmann, J. A. Whitehead, *Science* 256, 517 (1992).
2. K. A. Farley, J. H. Natland, H. Craig, *Earth Planet. Sci. Lett.* 111, 183 (1992).
3. H. Craig, *EOS* 71, 1669 (1990); *Caltech Plume Symposium Abstract Volume 1* (Plumacy Press, California Institute of Technology, Pasadena, CA 1991).

1 June 1992; accepted 17 August 1992

Response: In our paper on evidence for lower mantle plume entrainment, we stressed the potential role that He isotopes could play in supporting or rejecting our theory. We noted the lack of published He data on many key islands, so we welcome the comment by Farley and Craig.

Unfortunately, data are needed for He *and* Sr, Nd, and Pb isotopes on the same samples, and figure 1 of the comment by Farley and Craig does not help in this respect. Pitcairn shows enormous variations

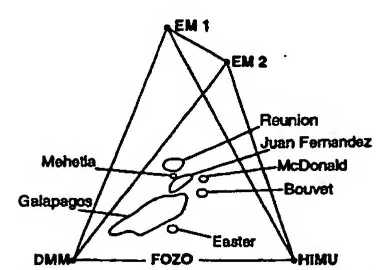


Fig. 1. Mantle tetrahedron from our report showing locations of additional high He 3/4 islands not shown in figure 1 of the comment by Farley and Craig. References for heavy isotope data may be found in (*3*); He data are from standard sources, Farley and Craig, and (*1*).